

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»**

**Інститут прикладного системного аналізу
Кафедра математичних методів системного аналізу**

«На правах рукопису»
УДК 519.233.2

«До захисту допущено»
Завідувач кафедри
_____ О.Л. Тимощук
«__» _____ 20__р.

Магістерська дисертація

на здобуття ступеня магістра

зі спеціальності 124 Системний аналіз

**на тему: «Оцінювання кредитоспроможності клієнтів банку на основі
регресійних моделей»**

Виконав:

студент II курсу, групи КА-61м
Сєдак Назар Денисович _____

Керівник:

доцент, к.ф. –м.н., доц.,
Каніовська І.Ю. _____

Рецензент:

доцент кафедри математичного аналізу
та теорії ймовірностей
НТУУ «КПІ ім. Ігоря Сікорського»
канд. фіз.-мат. наук Буценко Ю.П. _____

Засвідчую, що у цій магістерській
дисертації немає запозичень з праць
інших авторів без відповідних
посилань.
Студент _____

Київ 2018

ЗАВДАННЯ
на магістерську дисертацію студенту

Сєдаку Назару Денисовичу

1. Тема дисертації: «Оцінювання кредитоспроможності клієнтів банку на основі регресійних моделей», науковий керівник дисертації Каніовська Ірина Юріївна, канд. техн. наук, доцент.

затверджені наказом по університету від «27» березня 2018 р. №1028-с

2. Термін подання студентом дисертації: _____

3. Об'єкт дослідження: аналіз руху грошових потоків у визначенні кредитоспроможності.

4. Предмет дослідження: використання скорингової моделі як вхідних даних для регресійного аналізу.

5. Перелік завдань, які потрібно розробити: систематизувати існуючі методи математичні методи оцінювання кредитоспроможності позичальника, зібрати та обробити попередніх даних, побудова інтервалів для фінансових коефіцієнтів з відповідними значеннями для скорингової моделі, розробити модель для оцінки кредитоспроможності позичальників банку та реалізувати її програмно, провести експериментальні дослідження з використанням вибірки даних українського комерційного банку .

6. Орієнтовний перелік ілюстративного матеріалу: графіки, що характеризують характер впливу окремого фінансового коефіцієнта на кредитоспроможність позичальника в цілому, графік порівняння отриманих прогнозованих значень та реальних даних для звичайної та скорингової моделі.

7. Орієнтовний перелік публікацій: стаття «Improving the results of financial analysis by using scoring model during creditworthiness assessment», тези «Структурные особенности меметического алгоритма оптимизации прогноза при оценке кредитоспособности».

8. Дата видачі завдання:

Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Термін виконання етапів магістерської дисертації	Примітка
1	Грунтовне ознайомлення з предметною областю	16.03.2018	
2	Визначення структури магістерської дисертації; вивчення літератури, пошук додаткової літератури	27.03.2018	
3	Робота над першим розділом магістерської дисертації	04.04.2018	
4	Проведення наукового дослідження; робота над другим розділом магістерської дисертації	15.04.2018	
5	Проведення наукового дослідження; робота над статтею за результатами наукового дослідження	23.04.2018	
6	Робота над третім розділом магістерської дисертації; підготовка статті за результатами наукового дослідження; розроблення програмного забезпечення	30.04.2018	
7	Завершення роботи над основною частиною магістерської дисертації; робота над розділом з охорони праці	7.05.2018	
8	Оформлення текстової і графічної частин магістерської дисертації	15.05.2018	

Студент

Сєдак Н.Д.

Науковий керівник дисертації

Каніовська І.Ю.

РЕФЕРАТ

Магістерська дисертація: 104 с., 24 рис., 5 табл., 2 додатки, 35 джерел.

Об'єкт дослідження – методи та моделі оцінок кредитоспроможності клієнтів банку.

Предмет дослідження – регресійна модель на основі скорингової моделі.

Методи дослідження: методи теорії алгоритмів та програмування (для програмної реалізації розроблених алгоритмів); методи теорії ймовірності та математичної статистики (для побудови довірчих інтервалів скорингової моделі).

Метою дисертаційної роботи є побудова покращеної моделі для оцінювання здатності підприємства генерувати грошові кошти в обсягах і за строками здійснення необхідних платежів.

Для досягнення вказаної мети було розв'язано такі задачі:

- систематизувати існуючі методи математичні методи оцінювання кредитоспроможності позичальника;
- зібрати та обробити попередні данні;
- побудувати інтервали для фінансових коефіцієнтів з відповідними значеннями для скорингової моделі;
- розробити модель для оцінки кредитоспроможності позичальників банку та реалізувати її програмно;
- провести експериментальні дослідження з використанням вибірки даних українського комерційного банку.

КРЕДИТОСПРОМОЖНІСТЬ, СКОРИНГОВА МОДЕЛЬ, РЕГРЕСІЙНА МОДЕЛЬ, ФІНАНСОВИЙ АНАЛІЗ, ФІНАНСОВІ КОЕФІЦІЄНТИ, ПРОГНОЗ, ОЦІНКА, КОЕФІЦІЄНТ ПОКРИТТЯ ЗОБОВ'ЯЗАНЬ EBITDA.

ABSTRACT

The dissertation is performed on 103 sheets, it contains 2 applications and a list of references to used sources of 25 titles. In the work there are 10 figures and 32 tables.

Actuality of theme. The main purpose of the analysis of cash flow in determining the creditworthiness is to assess the ability of an enterprise to generate cash in volumes and on terms of making necessary payments.

The practice of crediting is carried out with the actions of a number of different and different risk factors, which may lead to the failure of the taxpayer at all times. At such a time, the bank is concerned with the study of credibility, thus, the study of the factors that may cause their failure. Meat and salvation of the report of the credit union are established in the knowledge of the health of the pupil at the beginning and in the last month, to ensure that the comradeship is commensurate with the degree of risk that the bank can take on the siege; the size of the keridite, which can be found in the dreams of the baptized ones, and in the hope of understanding.

As a result of this, there is a lack of efficiency in the way of banks, not only on the basis of the client's ability to access the data, but also to provide it with financial support for the worker. The objective factor of the financial position of the person and the risks of excessive risks in the various cases is to use the e-power to manage the credit resources and to receive revenues.

In the case of the americans, 35-40% of the prevailing problems arise in the analysis of the non-current financial statement of the financial strength of the post-election process at the previous stage of the program.

Relationship of work with scientific programs, plans, themes. The dissertation work was carried out in accordance with the research plan of the

department of applied mathematics of the National technical university of Ukraine "Kyiv polytechnic institute named after Igor Sikorsky".

The purpose and tasks of the study. The purpose of the dissertation work is to construct an improved model for assessing the ability of an enterprise to generate cash in volumes and on terms of making necessary payments.

To achieve this goal, the following tasks were solved:

To systematize existing methods of mathematical methods for the recognition of the credo of the creditor;

Collect and process preliminary data;

To construct inputs for the financial variables with the corresponding calculations for the scoring model;

To develop a model for the state of the creditworthiness and implement it programmatically;

Carry out pilot studies using a sample of ukrainian commercial bank data.

The object of the study is to analyze the flow of cash flows in determining the creditworthiness.

The subject of the study is the use of a scorecard model as input data for regression analysis.

Research methods. To solve the problem, the following methods were used: the analysis of objects research systems under uncertainty (for the development of a scoring model), optimization methods (for developing an improved method for solving the problem of credit assessment); methods of algorithm theory and programming (for software implementation of developed algorithms); methods of probability theory and mathematical statistics (for constructing confidence intervals of the scoring model).

Key words: creditworthiness, scoring model, regression model, financial analysis, financial coefficients, EBITDA, forecast, assessment.

ЗМІСТ

РОЗДІЛ 1.....	11
АНАЛІЗ ЗАДАЧІ ОЦІНЮВАННЯ КРЕДИТОСПРОМОЖНОСТІ ПОЗИЧАЛЬНИКА	11
1.1. Загальний огляд проблеми	11
1.2. Поняття про скоринг	19
1.3. Рекомендації Національного банку України	22
1.4. Рейтингові агентства	27
1.5. Оцінка кредитоспроможності юридичних осіб	32
1.6. Оцінка кредитоспроможності фізичних осіб	35
1.7. Формалізація задачі	36
РОЗДІЛ 2.....	38
МАТЕМАТИЧНІ МЕТОДИ ОЦІНЮВАННЯ КРЕДИТОСПРОМОЖНОСТІ ПОЗИЧАЛЬНИКА.....	38
2.1. Класична регресія	38
2.1.1. Загальні поняття.....	38
2.2. Логістична регресія	42
2.3. Розробка моделі	61
РОЗДІЛ 3.....	66
РЕАЛІЗАЦІЯ МОДЕЛІ	66
3.1. Опис вхідних даних	66
3.2. Попередня обробка даних	67
3.3. Побудова моделі	74
РОЗДІЛ 4.....	78
РОЗРОБКА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ.....	78

4.1. Функціональність програмного забезпечення	78
4.2. Опис інтерфейсу	79
РОЗДІЛ 5.....	85
РОЗРОБЛЕННЯ СТАРТАП ПРОЕКТУ	85
5.1 Опис ідеї.....	85
5.2 Технологічний аудит ідеї проекту	86
5.3 Аналіз ринкових можливостей запуску стартап-проекту....	87
5.4 Розроблення ринкової стратегії проекту	91
5.5 Розроблення маркетингової програми стартап-проекту	93
5.6 Висновки	95

ВСТУП

Основна мета аналізу руху грошових потоків у визначенні кредитоспроможності — оцінювання здатності підприємства генерувати грошові кошти в обсягах і за строками здійснення необхідних платежів.

Процес кредитування пов'язаний з діями численних і різноманітних факторів ризику, здатних спричинити непогашення позички у встановлений строк. Тому при наданні позичок банк займається вивченням кредитоспроможності, тобто вивченням факторів, які можуть викликати їхнє непогашення. Мета й завдання аналізу кредитоспроможності укладаються у визначенні здатності позичальника вчасно й у повному обсязі погасити заборгованість по позичці, ступінь ризику, що банк готовий взяти на себе; розміру кредиту, що може бути наданий у даних обставинах і, нарешті, умов його надання.

Зважаючи на це, виникає необхідність оцінки банком не тільки платоспроможності клієнта на певну дату, але й прогнозу його фінансової стабільності на перспективу. Об'єктивна оцінка фінансової стабільності позичальника й облік можливих ризиків по кредитних операціях дозволяють банку ефективно управляти кредитними ресурсами й отримувати прибуток.

За даними американських аналітиків 35-40% прострочених позичок виникає в результаті недостатньо глибокого аналізу фінансового становища позичальника на попередній стадії переговорів.

ФОРМАЛІЗАЦІЯ ЗАДАЧІ

Формалізуємо задачу використання математичних методів при оцінюванні кредитоспроможності позичальника. Маємо кредитний портфель з відомої історією кожного кредиту (повернення або неповернення). Необхідно визначити кредитоспроможність нового позичальника, від якого надходить запит. Введемо наступні позначення:

$X = (x_1, \dots, x_n)$ - вектор параметрів позичальника

Y - змінна, що характеризує кредитоспроможність позичальника або його дефолт (може бути або бінарною змінною, тоді 1 – означає неповернення кредиту (дефолт), 0 - повернення кредиту; або неперервною змінною з інтервалу $[0;1]$ і означати ймовірність дефолту позичальника – ймовірність неповернення ним кредиту).

Тоді кредитну історію портфелю кредитів, який складають m позичальників позначимо за допомогою пар (X_i, Y_i) , $i=1, 2, \dots, m$.

Типовими моделями, що можуть використовуватися для вирішення даної задачі (а, отже, і для побудови скорингових систем), є наступні:

- бальна система;
- множинна регресія;
- логістична регресія;
- кластерний аналіз;
- дискримінантний аналіз;
- дерева класифікації.

В даній роботі для побудови оцінки кредитоспроможності буде використана логістична регресія.

РОЗДІЛ 1

АНАЛІЗ ЗАДАЧІ ОЦІНЮВАННЯ

КРЕДИТОСПРОМОЖНОСТІ ПОЗИЧАЛЬНИКА

1.1. Загальний огляд проблеми

Процес кредитування пов'язаний з діями численних і різноманітних факторів ризику, здатних спричинити непогашення позички у встановлений строк. Тому при наданні позичок банк займається вивченням кредитоспроможності, тобто вивченням факторів, які можуть викликати їхнє непогашення. Мети й завдання аналізу кредитоспроможності укладаються у визначенні здатності позичальника вчасно й у повному обсязі погасити заборгованість по позичці, ступінь ризику, що банк готовий взяти на себе; розміру кредиту, що може бути наданий у даних обставинах і, нарешті, умов його надання.

Зважаючи на це, виникає необхідність оцінки банком не тільки платоспроможності клієнта на певну дату, але й прогнозу його фінансової стабільності на перспективу. Об'єктивна оцінка фінансової стабільності позичальника й облік можливих ризиків по кредитних операціях дозволяють банку ефективно управляти кредитними ресурсами й отримувати прибуток. [1]

За даними американських аналітиків 35-40% прострочених позичок виникає в результаті недостатньо глибокого аналізу фінансового становища позичальника на попередній стадії переговорів.

Клієнт, що звертається в банк за одержанням кредиту, повинен представити заявку, де міститимуться вихідні дані про необхідну позичку:

- мета;
- розмір кредиту;
- строк;

- передбачуване забезпечення.

До заявки повинні бути прикладені документи й фінансові звіти, що служать обґрунтуванням прохання про надання позики й пояснюючі причини звернення до банку. Ці документи необхідна складова частина заявки. Їхній ретельний аналіз проводиться на наступних етапах, після того, як представник банку проведе попереднє інтерв'ю з позичальником і зробить висновок про перспективність угоди. [2]

Ключовим моментом аналізу будь-якої заявки й супровідних документів, а також результатів бесід є визначення характеру позичальника і його кредитоспроможності. Кредитний інспектор, часто допомагаючи позичальникові готувати заявку, з'ясовує обсяг доходів і активів, що представляють цінність, (таких, як високоліквідні цінні папери або ощадні депозити), якими розташовує позичальник для погашення кредиту. Усні відповіді клієнта, як правило, містять значно більше інформації, чим відомості, викладені в письмовому виді.

Кредитні інспектори надають значення не тільки розміру, але й стабільності доходів позичальника (юридичних і фізичних осіб). Вони воліють одержувати від клієнта інформацію про чистий дохід (тобто доходи після всіх виплат і втримань). Якщо мова йде про індивідуального позичальника, то співробітники банку звичайно зв'язуються з роботодавцями клієнта для перевірки достовірності зазначеного ним рівня доходів і тривалості роботи в даній фірмі.

Непрямим показником розміру й стабільності доходу служать дані про середньоденний залишок на розрахунковому (депозитному) рахунку клієнта. Ці дані, представлені позичальником, інспектор зіставляє з інформацією, отриманої з відповідного банку.

Підтримка значних залишків на розрахункові (депозитних) рахунках у банках свідчить про надійність фінансового становища клієнта, його

фінансової дисциплінованості й серйозності наміру погасити одержуваний кредит.

Кредитні інспектори осудливо ставляться до появи так званої “піраміди боргу”, коли позичальник бере кредит в одній фірмі або банку для оплати боргу іншому кредиторіві. З поля зору інспектора не зникає значна й зростаюча заборгованість клієнта по кредитних картках і частому поверненні чеків, виписаних клієнтом з його рахунку. На основі подібних фактів робиться висновок про реальне фінансове становище клієнта і його вміння керувати коштами.

На цьому попередньому етапі роботи із кредитування інспектор зможе визначити порядність клієнта й достовірність наданої ним інформації. Після уважного вивчення всіх документів, наданих потенційним позичальником, проведення бесіди, оцінки інформації, отриманої у відповідь на питання, кредитний інспектор за узгодженням з керівництвом банку й начальником відділу ухвалює рішення щодо продовженні роботи із клієнтом або відмови йому. Якщо приймається рішення про продовження роботи із клієнтом, то інспектор комплектує кредитне досьє (включаючи в нього заявку й супровідні документи, відповіді на питання, записи бесід і т.д.) і здійснює ретельне вивчення кредитоспроможності клієнта. [3]

Перед ухваленням рішення про видачу кредиту банк повинен оцінити кредитоспроможність, платоспроможність і фінансову стабільність позичальника. При цьому відзначимо, що існують розходження у визначенні поняття “платоспроможність” і “кредитоспроможність”.

Платоспроможність позичальника - це його можливість і здатність вчасно погасити всі види зобов'язань і заборгованості. У той же час кредитоспроможність характеризується лише можливістю позичальника погасити кредиторську заборгованість. Разом з тим характеристика

кредитоспроможності повинна бути дещо іншою у порівнянні із платоспроможністю, оскільки погашення позичок можливе за рахунок виторгу від реалізації майна, прийнятого банком у заставу по позичці або завдяки використанню гарантії своєчасного повернення коштів або навіть за рахунок страхування погашення позички.

Застосовувані в цей час і рекомендовані способи оцінки кредитоспроможності спираються головним чином на аналіз даних про діяльність позичальника в попередньому періоді. Крім того, для оцінки можливостей позичальника погашати свої зобов'язання проводиться співставлення даних, наявних у звітності, про зобов'язання позичальника й джерелах коштів для їхнього погашення. Таке співставлення може робитися на основі даних про зобов'язання підприємства й джерелах коштів для їхнього погашення не тільки в цілому по сумах зобов'язань і джерелах погашення, але також по сумах зобов'язань і джерел їхнього погашення, що доводяться на окремі короткі періоди.

При різних варіантах оцінки кредитоспроможності їх поєднує те, що вони ґрунтуються на дані звітності й ставляться до попереднього періоду. При всьому значенні такої оцінки вона не може вичерпним образом характеризувати кредитоспроможність позичальника в майбутньому періоді.

При використанні головним чином даних звітності оцінка платоспроможності й фінансової стабільності позичальника розглядається лише як попередня.

Перевірки, здійснювані аудиторськими організаціями, використовуються для оцінки кредитоспроможності позичальника. Корисність таких перевірок знижується у зв'язку з тим, що в них використовуються дані звітності.

Основні недоліки оцінки кредитоспроможності позичальників пов'язані з тим, що вони ґрунтуються на даних звітності, що характеризують стан справ у попередньому періоді.

Доцільно використовувати закордонний досвід прогнозування фінансового стану позичальника в майбутньому періоді, для того, щоб приймати обґрунтоване рішення про надання позичок.

Значні розміри процентних ставок позначаються на збільшенні видатків позичальників, а відповідно й на їхній кредитоспроможності. Природно, це необхідно враховувати при оцінці кредитоспроможності. Стан платіжної дисципліни позичальника має важливе значення для характеристики його кредитоспроможності.

При оцінці кредитоспроможності підприємства варто розглядати неплатежі, у тому числі суми прострочених платежів - постачальникам, банку, фінансовим органам - як самостійну групу, а не розподіляти їх у різні види кредиторської заборгованості.

Підвищення рівня кредитоспроможності позичальників пов'язане з підвищенням їхньої відповідальності за своєчасне погашення позичок і разом з тим зниженням вимогливості банків при кредитуванні. Щодо цього заслуговує на увагу послідовне дотримання принципів кредитування, що з деяких пір незаслужено ігнорується.

Дотримання принципів кредитування розглядається в закордонній літературі як необхідна умова його доцільної організації. Так, у виданому в Англії посібнику з банківських послуг відзначається, що ключовим словом, у якому зосереджені вимоги при видачі позичок позичальникам, є термін "PARTS", що включає в себе:

- purpose - призначення, ціль
- amount - сума, розмір
- repayment - оплата, повернення (боргу й відсотків)
- term - строк

- security - забезпечення. застава

Існує безліч різних методик аналізу фінансового становища клієнта і його надійності з погляду своєчасного погашення боргу банку. У практиці банків США застосовуються “Правила шести “Сі”, у яких критерії відбору клієнтів позначені словами, що починаються буквами “Сі” (в основному, даного правила стосуються позичальників-юридичних осіб):

Шість основних принципів кредитування (правила шести “Сі”)

- character (характер)
- capacity (здатність)
- cash (кошти)
- collateral (забезпечення)
- conditions (умови)
- control (контроль)

1. Характер позичальника. Кредитний інспектор має бути переконаний у тому, що клієнт може досить точно вказати мету одержання кредиту, і має серйозні наміри погасити його. Якщо в інспектора немає належної впевненості щодо мети кредиту, що запитується, то вона повинна бути уточненою. Навіть у цьому випадку кредитному інспекторові слід установити: чи відповідально клієнт ставитися до позикових коштів, чи дає правдиві відповіді на питання банку й чи докладе всіх зусиль для виплати заборгованості за кредитом. Відповідальність, правдивість і серйозність намірів клієнта погасити всю заборгованість складають те, що кредитний інспектор називає характером позичальника.

Якщо в кредитного інспектора після вивчення документів, наданих позичальником, і бесід з ним виникли сумніви щодо надійності клієнта, його бажання й здатності погасити кредит, то клієнтові варто відмовити в кредитуванні. У протилежному випадку банк напевно буде мати справу із проблемним кредитом, погашення якого викликає серйозні сумніви.

2. Здатність запозичити кошти. Кредитний інспектор повинен бути впевнений у тім, що клієнт, що просить кредит, має юридичне право подавати кредитну заявку й підписувати кредитний договір. Дана характеристика клієнта відома як його здатність запозичити кредитні кошти. Наприклад, у більшості країн неповнолітні (тобто особи, що не досягли 18 років або 21 року) не мають право відповідати за кредитним договором. Банк зіштовхується із труднощами при стягненні такого кредиту. Кредитний інспектор повинен бути впевнений у тім, що керівник або представник компанії (банку), що звертається за кредитом, має відповідні повноваження, надані йому засновниками або радою директорів, на проведення переговорів і підписання кредитного договору від імені компанії (банку) для того, щоб визначити, які особи вповноважені на підписання кредитного договору. Стягнення по суду коштів за кредитним договором, підписаному не уповноваженими на те особами, може виявитися неможливим, і банк понесе значні збитки.

3. Кошти. Ключовий момент будь-якої кредитної заявки складається у визначенні можливостей позичальника погасити кредит. У цілому позичальник має тільки три джерела погашення отриманих їм кредитів:

- а) потоки готівки;
- б) продаж або ліквідація активів;
- в) залучення фінансів.

Кожної із зазначених джерел може забезпечити залишкову суму коштів для погашення кредиту. Однак банкіри воліють вибирати як основне джерело погашення кредиту позичальником потік готівки, оскільки продаж активів може погіршити баланс позичальника, а його додаткові запозичення можуть послабити позиції банку як кредитора. Недостатність потоку готівки є важливим показником погіршення фінансового стану фірми й взаємин із кредиторами.

Одним з переваг даної формули є те, що за її допомоги кредитний інспектор банку може визначити ті сторони діяльності клієнта, які відображають кваліфікацію й досвід його менеджерів, а так само стан ринку, в умовах якого працює клієнт. Позичальник, що працює завдяки переважному використанню комерційного кредиту (кредиторської заборгованості), стане для банку проблемним клієнтом. Більшість банків будуть випробовувати сумніви щодо доцільності надання йому кредиту. [4]

4. Забезпечення. При оцінці забезпечення за кредитною заявкою кредитний інспектор повинен одержати відповідь на питання: чи розташовує позичальник достатнім капіталом або якісними активами для надання необхідного забезпечення по кредиту? Кредитний інспектор звертає особливу увагу на такі характеристики, як: термін служби, стан і структура активів позичальника. Якщо активи позичальника - це застаріле обладнання й технологія, то їхня цінність як кредитне забезпечення невелике, оскільки подібні активи буде важко перетворити в наявні кошти у випадку недостатності доходів позичальника для погашення заборгованості за кредитом.

5. Умови. Кредитний інспектор повинен знати, стан справ позичальника або положення, що складається у відповідній галузі, а також те, як зміна економічних і інших умов може вплинути на процес погашення кредиту. За документами кредит може здатися надійним з погляду забезпечення, але ступінь його надійності може понизитися в результаті скорочення обсягу продажів або доходу в умовах економічного спаду або росту процентних ставок, викликаного інфляцією. Для оцінки стану галузі й економічних умов більшість банків створюють інформаційні центри з базою даних, збирають різні інформаційні матеріали й підсумкові документи про наукові дослідження з галузей, у яких діють їхні основні позичальники.

6. Контроль. Останнім фактором оцінки кредитоспроможності позичальника виступає контроль, що зводиться до отримання відповідей на такі питання, як: наскільки зміна законодавства, правової, економічної й політичної обстановки може негативно вплинути на діяльність позичальника і його кредитоспроможність?

Схема показників, названа “Правила шести “Сі” і застосовувана банками США, один з багатьох підходів оцінки кредитоспроможності й надійності позичальників. Банки країн з розвиненою ринковою економікою застосовують складну систему великої кількості показників для оцінки кредитоспроможності клієнтів. Ця система диференційована залежно від характеру позичальника (фірма, частка особа, вид діяльності), так само може ґрунтуватися як на сальдових, так і оборотних показниках звітності клієнта. Так ряд американських економістів описує систему оцінки кредитоспроможності, побудовану на сальдових показниках звітності.

1.2.Поняття про скоринг

У світовій практиці існує два основних методи оцінки ризику кредитування, які можуть застосовуватися як окремо, так і в сполученні один з одним:

- суб’єктивний висновок експертів або кредитних інспекторів;
- автоматизовані системи скорингу.

Скоринг є одним з найбільш успішних прикладів використання математичних і статистичних методів у бізнесі.

Скоринг являє собою математичну або статистичну модель, за допомогою якої на основі кредитної історії «минулих» клієнтів банк намагається визначити, наскільки велика ймовірність, що конкретний потенційний позичальник поверне кредит у строк.

У західній банківській системі, коли позичальник звертається за кредитом, банк може мати наступну інформацію для аналізу:

- анкета, що заповнює позичальник;
- інформація на даного позичальника із кредитного бюро - організації, у якій зберігається кредитна історія всього дорослого населення країни;
- дані рухів по рахунках, якщо мова йде про вже діючого клієнта банку.

Кредитні аналітики оперують наступними поняттями: «характеристики» клієнтів (у математичній термінології - змінні, фактори) і «ознаки» - значення, які приймає змінна. Якщо уявити собі анкету, що заповнює клієнт, то характеристиками є питання анкети (вік, родинний стан, професія), а ознаками - відповіді на ці питання. [5]

У самому спрощеному виді скорингова модель являє собою зважену суму певних характеристик. У результаті отримується інтегральний показник; чим він вище, тим вище надійність клієнта, і банк може впорядкувати своїх клієнтів по ступені зростання кредитоспроможності.

Інтегральний показник кожного клієнта порівнюється з певним числовим порогом, або лінією поділу, що, по суті, є лінією беззбитковості й розраховується з відношення, скільки в середньому потрібно клієнтів, які платять у строк, для того, щоб компенсувати збитки від одного боржника. Клієнтам з інтегральним показником вище цієї лінії видається кредит, клієнтам з інтегральним показником нижче цієї лінії – не видається.

Але в скорингу існує дві основні проблеми. Перша полягає в тому, що класифікація вибірки проводиться тільки на клієнтах, яким надали кредит. Ми ніколи не довідаємося, як би повелися клієнти, яким у кредиті було відмовлено: цілком можливо, що якась частина виявилася б цілком прийнятними позичальниками.

Але, як правило, прийняття рішення про відмову в кредиті приймається на підставі досить серйозних причин. Банки фіксують ці причини відмови й зберігають інформацію про клієнтів, яким було відмовлено. Це дозволяє їм відновлювати первісну популяцію клієнтів, що зверталися за кредитом.

Друга проблема полягає в тому, що люди із часом змінюються, змінюються й соціально-економічні умови, що впливають на поведінку людей. Тому скорингові моделі необхідно розробляти на вибірці з найбільш «свіжих» клієнтів, періодично перевіряти якість роботи системи й, коли якість погіршується, розробляти нову модель. На Заході нова модель розробляється в середньому раз у півтора року, період між заміною моделі може варіюватися залежно від того, наскільки стабільної була економіка в цей час.

Несприятливим фактором розвитку скорингу у нашій країні є відсутність широко-функціонуючих кредитних бюро. Але, з іншого боку, на Заході існує проблема перевірки вірогідності інформації, що людина вказує про себе в анкеті. В Україні більша частина такої інформації знаходиться в паспорті. Банкам досить мати паспортної даної й дані трудової книжки - от і вихідний матеріал для аналізу.

Ще один несприятливий фактор - недостатня поширеність універсальних статистичних пакетів, що ускладнює проведення математичного аналізу даних позичальників.

На Заході при кредитуванні юридичних осіб скоринг-моделі поширені не настільки широко, як у споживчому кредиті. Це пов'язане з тим, що для розробки моделі дуже важко набрати достатню кількість компаній, подібних один з одним: компанії сильно відрізняються по розмірі, обороту, секторам економіки. Чим крупніше підприємство, тим складніше підібрати аналогічні підприємства для порівняння.

В останні роки значні зрушення відбулися в розробці скоринг-моделей для малого бізнесу. Застосування скоринга для малого й середнього бізнесу виявилось можливим саме в зв'язку великої кількості подібних між собою підприємств.

Потрібно відзначити, що в Україні впровадження скоринга гальмується не стільки об'єктивними, скільки суб'єктивними причинами, пов'язаними з недовірливим відношенням банківських менеджерів до математичних і статистичних методів. Для аналізу своїх клієнтів потрібно відносно небагато - кредитна історія минулих клієнтів і статистичний пакет, - а віддача може бути значною. Серед переваг скорингових систем західні банкіри вказують, у першу чергу, зниження рівня неповернення кредиту. Далі відзначається швидкість і неупередженість у прийнятті рішень, можливість ефективного керування кредитним портфелем, відсутність необхідності тривалого навчання персоналу.

Отже, скоринг являє собою автоматизовані системи оцінки кредитного ризику, які широко використовуються в США й Західній Європі. Як вихідний матеріал для скоринга використовується різноманітна інформація про минулих клієнтів, на основі якої за допомогою різних статистичних і нестатистичних методів класифікації робиться прогноз про кредитоспроможність майбутніх позичальників. Скорингові системи дозволяють банківським працівникам швидко ухвалювати рішення щодо кредитуванні, регулювати обсяги кредитування залежно від ситуації на ринку й визначати оптимальне співвідношення між прибутковістю кредитних операцій і рівнем ризику.

1.3. Рекомендації Національного банку України

Основним документом НБУ, що регулює порядок проведення оцінки кредитоспроможності позичальника та вимоги до даної оцінки є „Положення про порядок формування та використання резерву для

відшкодування можливих втрат за кредитними операціями банків” (надалі – Положення).

Відповідно до даного Положення критерії оцінки фінансового стану позичальника встановлюються кожним банком самостійно його внутрішніми положеннями щодо проведення активних операцій (кредитних) та методикою проведення оцінки фінансового стану позичальника (контрагента банку) з урахуванням вимог Положення, у яких мають бути визначені ґрунтовні, технічно виважені критерії економічної оцінки фінансової діяльності позичальників (контрагентів банку) на підставі аналізу їх балансів і звітів про фінансові результати в динаміці тощо. Методика проведення оцінки фінансового стану позичальника (контрагента банку), яка розроблена банком, є невід'ємним додатком до внутрішньобанківського положення банку про кредитування. [6]

Оцінку фінансового стану позичальника/контрагента банку з урахуванням поточного стану обслуговування позичальником/контрагентом кредитної заборгованості банк здійснює кожного разу під час укладання договору про здійснення кредитної операції, а надалі для:

- банків - не рідше ніж один раз на місяць;
- інших юридичних осіб - не рідше ніж один раз на три місяці;
- фізичних осіб - періодичність оцінки їх фінансового стану визначається банками самостійно з урахуванням стану обслуговування боргу та строковості кредиту, але не рідше ніж один раз у рік (або за результатами фінансового року). Якщо обслуговування кредитної заборгованості відбувається із простроченням або пролонгацією, то оцінка фінансового стану позичальника - фізичної особи має відбуватися постійно (щомісяця або щокварталу) протягом періоду несвоєчасного погашення боргу -

крім випадків, пов'язаних з поважними причинами (документально підтверджені факти відрадження, хвороба тощо).

Банк перевіряє кредитоспроможність позичальника, заборгованість за кредитом якого включена до портфеля однорідних споживчих кредитів, за власною методикою.

Банк самостійно встановлює періодичність перегляду кредитоспроможності позичальника за кредитами, заборгованість за якими включена до портфеля однорідних споживчих кредитів.

Результати оцінки фінансового стану позичальників/контрагентів банку мають зберігатися в банку відповідно до строків зберігання документів з кредитування, встановлених законодавством України, особливо щодо великих кредитів, кредитів пов'язаними з банками (інсайдерами).

Зазначені в цьому Положенні вимоги щодо оцінки фінансового стану позичальника є мінімально необхідними. Банки мають право самостійно встановлювати додаткові критерії оцінки фінансового стану позичальника, що підвищують вимоги до показників з метою адекватної оцінки кредитних ризиків та належного контролю за ними.

Банки самостійно встановлюють нормативні значення та відповідні бали для кожного показника залежно від його вагомості (значимості) серед інших показників, що можуть свідчити про найбільшу ймовірність виконання позичальником (контрагентом банку) зобов'язань за кредитними операціями.

Вагомість кожного показника визначається індивідуально для кожної групи позичальників (контрагентів банку) залежно від кредитної політики банку, особливостей клієнта (галузь економіки, сезонність виробництва, обіговість коштів тощо), ліквідності балансу, становища на ринку тощо.

Клас позичальника (контрагента банку) за результатами оцінки його фінансового стану визначається на підставі основних показників та коригується з урахуванням додаткових (суб'єктивних) показників.

Класифікація позичальників - юридичних осіб (у тому числі банків) здійснюється за результатами оцінки їх фінансового стану:

Клас "А" - фінансова діяльність добра, що свідчить про можливість своєчасного виконання зобов'язань за кредитними операціями, зокрема погашення основної суми боргу та відсотків за ним відповідно до умов кредитної угоди; економічні показники в межах установлених значень (відповідно до методики оцінки фінансового стану позичальника, затвердженої внутрішніми документами банку); вище керівництво позичальника має відмінну ділову репутацію; кредитна історія позичальника - бездоганна; крім того, позичальники-банки (резиденти) і банки-нерезиденти, що зареєстровані в країнах - членах СНД, дотримуються економічних нормативів. Одночасно можна зробити висновок, що фінансова діяльність і надалі проводитиметься на високому рівні. До цього класу можуть належати інші позичальники-банки (нерезиденти), що мають кредитний рейтинг не нижче ніж показник А, підтверджений у бюлетені однієї з провідних світових рейтингових компаній (Fitch IBCA, Standard & Poor's, Moody's тощо). [7]

Клас "Б" - фінансова діяльність позичальника цієї категорії близька за характеристиками до класу "А", але ймовірність підтримування її на цьому рівні протягом тривалого часу є низькою. Позичальники/контрагенти банку, які належать до цього класу, потребують більшої уваги через потенційні недоліки, що ставлять під загрозу достатність надходжень коштів для обслуговування боргу та стабільність одержання позитивного фінансового результату їх діяльності; крім того, позичальники-банки (резиденти) і банки-нерезиденти, що зареєстровані в країнах - членах СНД, дотримуються економічних

нормативів. Аналіз коефіцієнтів фінансового стану позичальника може свідчити про негативні тенденції в діяльності позичальника. Недоліки в діяльності позичальників, які належать до класу "Б", мають бути лише потенційними. За наявності реальних недоліків клас позичальника потрібно знизити. До цього класу можуть належати інші позичальники-банки (нерезиденти), що мають кредитний рейтинг не нижче ніж "інвестиційний клас", що підтверджений у бюлетені однієї з провідних світових рейтингових компаній (Fitch IBCA, Standard & Poor's, Moody's тощо).

Клас "В" - фінансова діяльність задовільна і потребує більш детального контролю, крім того, позичальники - банки (резиденти) і банки країн - членів СНД дотримуються економічних нормативів. Надходження коштів і платоспроможність позичальника свідчать про ймовірність своєчасного погашення кредитної заборгованості в повній сумі та в строки, передбачені договором, якщо недоліки не будуть усунені. Одночасно спостерігається можливість виправлення ситуації і покращання фінансового стану позичальника. Забезпечення кредитної операції має бути ліквідним і не викликати сумнівів щодо оцінки його вартості, правильності оформлення угод про забезпечення тощо. До цього класу можуть належати позичальники-банки (нерезиденти), що мають кредитний рейтинг не нижче ніж показник В, підтверджений у бюлетені однієї з провідних світових рейтингових компаній (Fitch IBCA, Standard & Poor's, Moody's тощо).

Клас "Г" - фінансова діяльність незадовільна (економічні показники не відповідають установленим значенням) і спостерігається її нестабільність протягом року; є високий ризик значних збитків; ймовірність повного погашення кредитної заборгованості та відсотків/комісій за нею є низькою; проблеми можуть стосуватися стану забезпечення за кредитом, потрібної документації щодо забезпечення, яка

свідчить про наявність (схоронність) і його ліквідність тощо. Якщо під час проведення наступної класифікації немає безсумнівних підтверджень поліпшити протягом одного місяця фінансовий стан позичальника банку або рівень забезпечення за кредитною операцією, то його потрібно класифікувати на клас нижче (клас "Д"). До цього класу належить позичальник/контрагент банку, проти якого порушено справу про банкрутство.

Клас "Д" - фінансова діяльність незадовільна і є збитковою; показники не відповідають установленим значенням, кредитна операція не забезпечена ліквідною заставою (або безумовною гарантією), ймовірності виконання зобов'язань позичальником/контрагентом банку практично немає. До цього класу належить позичальник/контрагент банку, що визнаний банкрутом в установленому чинним законодавством порядку.

За результатами оцінки фінансового стану позичальник зараховується до відповідного класу.

Позичальник/контрагент банку, проти якого порушено справу про банкрутство, а також банки, в яких відкликано банківську ліцензію, не можуть бути віднесені до класу, вище ніж клас "Г".

Якщо банк здійснює довгострокове кредитування під інвестиційний проект, розрахунок ефективності якого (у тому числі доходу, що планується отримати від його реалізації) забезпечує погашення кредиту і відсотків/комісій за ним відповідно до умов кредитної угоди, а також за умови, що бізнес-план реалізації інвестиційного проекту виконується, то клас позичальника може бути підвищено на один рівень.

1.4. Рейтингові агентства

Одним з методів отримання додаткової інформації про кредитоспроможність позичальника є використання інформації

рейтингового агентства. Рейтингове агентство - комерційна організація, що займається оцінкою платоспроможності емітентів, боргових зобов'язань, якості корпоративного керування, якості керування активами й т.п. Найбільш відомий продукт рейтингових агентств - це оцінка платоспроможності - кредитний рейтинг. Він відображає ризик невиконання по борговому зобов'язанню й впливає на величину відсоткової ставки, на вартість і доходність зобов'язань/активів. При цьому більш високий рейтинг відповідає меншому ризику невиконання. [8]

У світі налічується більше 100 рейтингових агентств. До найбільш відомих міжнародних агентств відносяться:

- Fitch;
- Moody's;
- Standard & Poor's.

Можна побудувати зведену шкалу рейтингів найбільш відомих кредитних агентств:

Таблиця 1.4.1. Зведена шкала рейтингів

Fitch	Moody 's	Standard & Poor`s
AAA	Aaa	AAA
AA+, AA, AA-	Aa1, Aa2, Aa3	AA+, AA, AA-
A+, A, A- A-	A1, A2, A3	A+, A, A- A-
BBB+	Baa1	BBB+
BBB	Baa2	BBB
BBB-	Baa3	BBB-
BB+	Ba1	BB+
BB	Ba2	BB
BB-	Ba3	BB-
B+, B, B-B-	B1, B2, B3	B+, B, B-B-

CCC+, CCC, CCC-	Сaa, Са	CCC+, CCC, CCC-
D	C	CC, C, D

Розглянемо докладніше шкалу рейтингів Standard & Poor's. У якості міжнародного рейтингового агентства Standard & Poor's займається присвоєнням короткострокових і довгострокових кредитних рейтингів як емітентам, так і окремим борговим зобов'язанням.

Міжнародна шкала кредитних рейтингів Standard & Poor's служить для задоволення потреб учасників глобальних (міжнародних) фінансових ринків. Оцінки по цій шкалі дозволяють порівнювати між собою надійність емітентів і зобов'язань різних держав.

Довгострокові рейтинги оцінюють здатність емітента вчасно виконувати свої боргові зобов'язання. Рейтингові оцінки, що виставляються компанією, мають літерне позначення: від оцінки AAA, що привласнюється винятково надійним емітентам, до оцінки D, що привласнюється емітенту, що оголосив дефолт. Між оцінками AA і B можуть бути проміжні оцінки, позначувані знаками плюс і мінус (наприклад, BBB+, BBB і BBB-).

Рейтинги інвестиційного класу:

- AAA - емітент має винятково високі можливості по виплаті відсотків по боргових зобов'язаннях і самих боргах.
- AA - емітент має дуже високі можливості по виплаті відсотків по боргових зобов'язаннях і самих боргах.
- A - можливості емітента по виплаті відсотків і боргів оцінюються високо, але залежать від економічної ситуації.
- BBB - платоспроможність емітента вважається задовільною.

Рейтинги спекулятивного класу:

- BB - емітент платоспроможний, але несприятливі економічні умови можуть негативно вплинути на можливості виплат.

- В - емітент платоспроможний, але несприятливі економічні умови найімовірніше негативно вплинуть на його можливості й готовність проводити виплати по боргах.
- CCC - емітент зазнає труднощів з виплатами до борговим зобов'язанням і його можливості залежать від сприятливих економічних умов.
- CC - емітент зазнає серйозних труднощів з виплатами по боргових зобов'язаннях.
- C - емітент зазнає серйозних труднощів з виплатами по боргових зобов'язаннях, можливо була ініційована процедура банкрутства, але виплати по боргових зобов'язаннях усе ще провадяться.
- SD - емітент відмовився від виплат по деяких зобов'язаннях.
- D - був оголошений дефолт і S&P думає, що емітент відмовиться від виплат по більшості або по всіх зобов'язаннях.
- NR - рейтинг не привласнений.

Короткострокові рейтинги оцінюють імовірність своєчасного погашення короткострокових боргових зобов'язань. Кредитні рейтинги по короткострокових боргових зобов'язаннях, що виставляються Standard & Poor's, мають літерно-цифрове позначення: від найвищої оцінки А-1 до найнижчої оцінки D. Більше надійні зобов'язання з категорії А-1 можуть бути позначені знаком плюс. Оцінки з категорії В також можуть бути уточнені цифрою (В-1, В-2, В-3).

- А-1 - емітент має винятково високі можливості по погашенню даного боргового зобов'язання.
- А-2 - емітент має високі можливості по погашенню даного боргового зобов'язання, але ці можливості більш чутливі до несприятливих економічних умов.
- А-3 - несприятливі економічні умови цілком імовірно послабляють можливості емітента по погашенню даного боргового зобов'язання.

- В - боргове зобов'язання має спекулятивний характер. Емітент має можливості по його погашенню, але ці можливості дуже чутливі до несприятливих економічних умов.
- С - можливості емітента по погашенню даного боргового зобов'язання обмежені й залежать від наявності сприятливих економічних умов.
- D - по даному короткостроковому борговому зобов'язанню був оголошений дефолт.

Поряд з міжнародною шкалою кредитного рейтингу Standard & Poor's підтримує так само ряд національних шкал, у тому числі й українську. Національні шкали призначені для задоволення потреб учасників національних фінансових ринків. Рейтинг емітента й рейтинг боргового зобов'язання по національній шкалі відображають оцінку відносної надійності емітентів і боргових зобов'язань, що є присутнім на національному ринку. Національна шкала надає більше можливостей для розрізнення кредитоспроможності емітентів, тому що виключає деякі суверенні ризики, зокрема ризик переводу коштів за межі держави й інші систематичні ризики, рівною мірою характерні для всіх емітентів на даному ринку.

Оскільки рейтинги по національній шкалі відображають національну специфіку, порівнювати рейтинги по різних національних шкалах не має змісту. Точно так само не порівнянні рейтинги по національній шкалі й по міжнародній шкалі.

Шкала кредитного рейтингу Standard & Poor's для України використовує традиційні символи Standard & Poor's із префіксом «ua».

Поряд із присвоєнням рейтингу S&P указує також прогноз зміни рейтингу в найближчі два-три роки:

- Позитивний прогноз - можливе підвищення рейтингу.
- Негативний прогноз - можливе зниження рейтингу.

- Стабільний прогноз - рейтинг швидше за все залишиться незмінним.
- Прогноз, що розвивається, - можливо як підвищення так і зниження рейтингу.

1.5. Оцінка кредитоспроможності юридичних осіб

Для здійснення оцінки кредитоспроможності позичальника - юридичної особи банк має враховувати такі основні економічні показники його діяльності:

- платоспроможність (коефіцієнти миттєвої, поточної та загальної ліквідності);
- фінансова стійкість (коефіцієнти маневреності власних коштів, співвідношення залучених і власних коштів);
- обсяг реалізації;
- обороти за рахунками (співвідношення надходжень на рахунки позичальника і суми кредиту, наявність рахунків в інших банках; наявність картотеки неплатежів);
- склад та динаміка дебіторсько-кредиторської заборгованості (за останній звітний та поточний роки);
- собівартість продукції;
- прибутки та збитки;
- рентабельність;
- кредитна історія (погашення кредитної заборгованості в минулому, наявність діючих кредитів).

Банки повинні визначати значення показників платоспроможності позичальника та його фінансової стійкості з урахуванням статистичних даних галузі господарства, у якій він працює, і даних про результати його діяльності.

1. Платоспроможність позичальника може бути визначена за такими показниками:

- коефіцієнт миттєвої ліквідності (КЛ1), що характеризує те, як швидко короткострокові зобов'язання можуть бути погашені високоліквідними активами:

$$KL = \frac{A_{\text{в}}}{Z_{\text{п}}} \quad (1.4.1)$$

де $A_{\text{в}}$ - високоліквідні активи, до яких належать грошові кошти, їх еквіваленти та поточні фінансові інвестиції,

$Z_{\text{п}}$ - поточні (короткострокові) зобов'язання, що складаються з короткострокових кредитів і розрахунків з кредиторами.

- коефіцієнт поточної ліквідності (КЛ2), що характеризує можливість погашення короткострокових зобов'язань у встановлені строки:

$$KL = \frac{A_{\text{л}}}{Z_{\text{п}}} \quad (1.4.2)$$

де $A_{\text{л}}$ - ліквідні активи, що складаються з високоліквідних активів, дебіторської заборгованості, векселів одержаних,

$Z_{\text{п}}$ - поточні (короткострокові) зобов'язання, що складаються з короткострокових кредитів і розрахунків з кредиторами.

- коефіцієнт загальної ліквідності (КП), що характеризує те, наскільки обсяг короткострокових зобов'язань і розрахунків можна погасити за рахунок усіх ліквідних активів:

$$KL = \frac{A_{\text{о}}}{Z_{\text{п}}} \quad (1.4.3)$$

де $A_{\text{о}}$ - оборотні активи,

$Z_{\text{п}}$ - поточні (короткострокові) зобов'язання, що складаються з короткострокових кредитів і розрахунків з кредиторами.

2. Фінансова стійкість позичальника визначається за такими показниками:

- коефіцієнт маневреності власних коштів (КМ), що характеризує ступінь мобільності використання власних коштів:

$$KM = \frac{B_k - A_n}{B_k} \quad (1.4.4)$$

де B_k - власний капітал підприємства,

A_n - необоротні активи.

- коефіцієнт незалежності (КН), що характеризує ступінь фінансового ризику:

$$KH = \frac{Z_k}{B_k} \quad (1.4.5)$$

де Z_k - залучені кошти (довгострокові та поточні зобов'язання),

B_k - власний капітал.

3. Рентабельність позичальника визначається за такими показниками:

- рентабельність активів:

$$PA = \frac{Пч}{A} \quad (1.4.6)$$

де $Пч$ - чистий прибуток,

A - активи;

- рентабельність продажу:

$$PP = \frac{Пч}{Op} \quad (1.4.7)$$

де $Пч$ - чистий прибуток,

Op - обсяг реалізації продукції (без ПДВ).

Також можуть бути враховані суб'єктивні чинники, що характеризуються такими показниками:

- ринкова позиція позичальника та його залежність від циклічних і структурних змін в економіці та галузі промисловості;
- конкуренція у сфері діяльності;

- технологічний рівень позичальника;
- наявність державних замовлень і державна підтримка позичальника;
- ефективність управління позичальника;
- професіоналізм керівництва та його ділова репутація;
- акціонери позичальника;
- ділова репутація;
- банківська та кредитна історія позичальника.

1.6. Оцінка кредитоспроможності фізичних осіб

На практиці банки найчастіше застосовують скорингові системи для здійснення оцінки кредитоспроможності фізичних осіб, особливо для споживчого кредитування. Кредитування юридичних осіб є складнішим процесом, там фігурують більші розміри кредитів, тому потрібен більш індивідуальний підхід.

Набільш застосованими в якості параметрів скорингових систем наступні показники позичальників-фізичних осіб:

- вік;
- стать;
- освіта;
- місце та термін проживання;
- сімейний стан;
- кількість дітей;
- власність;
- кредитна історія;
- місце та досвід роботи;
- показники платоспроможності тощо.

Приділимо увагу показникам платоспроможності. Основними з них є наступні:

$$PTI = \frac{\text{виплати за кредитом}}{\text{сукупний дохід}} * 100\%$$

$$OTI = \frac{\text{регулярні витрати позичальника}}{\text{сукупний дохід}} * 100\%$$

Враховується місячне значення кожного з параметрів (виплати за кредитом, сукупний дохід тощо) цих показників. Показники платоспроможності можуть не виступати в якості параметрів скорингової моделі, але їх обчислення є необхідним для визначення фізичної можливості позичальника сплачувати кредит (дані показники мають бути менше 100%).

1.7. Формалізація задачі

Формалізуємо задачу використання математичних методів при оцінюванні кредитоспроможності позичальника. Маємо кредитний портфель з відомої історією кожного кредиту (повернення або неповернення). Необхідно визначити кредитоспроможність нового позичальника, від якого надходить запит. Введемо наступні позначення:

$X = (x_1, \dots, x_n)$ - вектор параметрів позичальника

Y - змінна, що характеризує кредитоспроможність позичальника або його дефолт (може бути або бінарною змінною, тоді 1 – означає неповернення кредиту (дефолт), 0 - повернення кредиту; або неперервною змінною з інтервалу $[0;1]$ і означати ймовірність дефолту позичальника – ймовірність неповернення ним кредиту).

Тоді кредитну історію портфелю кредитів, який складають m позичальників позначимо за допомогою пар (X_i, Y_i) , $i=1, 2, \dots, m$.

Типовими моделями, що можуть використовуватися для вирішення даної задачі (а, отже, і для побудови скорингових систем), є наступні:

- Бальна система;
- Множинна регресія;

- Логістична регресія;
- Кластерний аналіз;
- Дискримінантний аналіз;
- Дерева класифікації.

В даній роботі для побудови оцінки кредитоспроможності буде використана логістична регресія.

РОЗДІЛ 2

МАТЕМАТИЧНІ МЕТОДИ ОЦІНЮВАННЯ КРЕДИТОСПРОМОЖНОСТІ ПОЗИЧАЛЬНИКА

2.1. Класична регресія

2.1.1. Загальні поняття

Регресійна модель описує об'єктивно існуючі між явищами кореляційні зв'язки. За своїм характером кореляційні зв'язки надзвичайно складні та різноманітні. В одних випадках результат у зі зміною фактора x зростає чи зменшується рівномірно, в інших – нерівномірно. Іноді зростання може змінитися зменшенням і навпаки. Простежити всі ці взаємозв'язки і встановити точний функціональний вид практично неможливо. А тому при виборі типу функції йдеться лише про апроксимацію відносно простими функціями незрівнянно більш складних за своєю природою взаємозв'язків. На практиці перевагу віддають моделям, які є лінійними або приводяться до лінійного виду шляхом перетворення змінних, наприклад логарифмування. Такий підхід, безперечно, містить у собі певну умовність, оскільки передбачає однаковий характер зв'язку з усіма факторами. Проте використання надто складних функцій неминуче веде до збільшення кількості параметрів, а отже, зменшує точність вимірювання та ускладнює інтерпретацію результатів.

При обґрунтуванні типу функцій слід враховувати й той факт, що межі варіації корельованих ознак у конкретних умовах простору і часу, в конкретній сукупності значно вужчі за їх можливі значення, і в цих межах варіації навіть лінійна функція може задовільно апроксимувати зв'язок.

У лінійному щодо параметрів рівняння регресії індивідуальне значення результативного показника y_j (де j – порядковий номер одиниці сукупності) записується так:

$$y_j = b_0 + \sum_{i=1}^m b_i x_i + e_j ,$$

де b_0 – вільний член рівняння; економічного зросту, як правило, не має, лише окреслює область існування моделі;

b_i – коефіцієнт регресії; показує, як в середньому змінюється y зі зміною x_i на одиницю її шкали вимірювання за незмінності інших включень в модель факторів і за інших рівних умов;

$e_j = y_j - Y_j$ – залишкова величина.

У регресійній моделі основне навантаження покладається на коефіцієнт регресії b_i , він розглядається як своєрідна міра „очищеного” впливу x_i на y і називається ефектом впливу. Процедура оцінювання параметрів регресійної моделі ґрунтується на методі найменших квадратів (МНК). Наведемо загальну схему розрахунку статистичних характеристик моделі, акцентуючи увагу на їх змістовній інтерпретації.

Первинна інформація представляється як матриця факторних ознак X розміром $(n \cdot m)$ і вектора результативної ознаки y розміром $(n \cdot 1)$. Задля зручності використання алгоритмів МНК матриця X розширюється за рахунок додатково введеної фіктивної змінної x_0 , вектор якої представлений одиницями. Параметри моделі – вектор $B = [b_0, b_1, b_2, \dots, b_m]$ визначаються розв’язуванням системи нормальних рівнянь, яка записується так:

$X'XB = X'y$, де $X'X$ – матриця розміром $n(m+1)$.

Послідовність розрахунків включає етапи:

- обчислення матриці $X'X$ і вектора $X'y$;
- обертання матриці $C = (X'X)^{-1}$;
- розрахунок параметрів $B = CX'y$;

- визначення теоретичних значень результативної ознаки

$$Y = \sum_{i=1}^m b_s x_i \text{ та залишків } e_j = y_j - Y_j.$$

Значення коефіцієнтів регресії певною мірою залежать від складу введених у модель факторів. З розширенням ознакової множини моделі відбувається перерозподіл попередньо введених факторів. Чим вагоміший вплив нововведеного фактора, тим помітніші зміни. [9]

2.1.2. Адекватність моделі

Адекватність регресійної моделі означає здатність її правильно описати реальну структуру взаємозв'язків між ознаками x_i та y . Методологічною основою вирішення проблеми адекватності є теоретичний, змістовний аналіз матеріальної природи процесу (явища) та обґрунтування типу й структури моделі, яка описує механізм його формування. Практично з метою забезпечення адекватності моделі змістовний аналіз поєднується з формальними процедурами перевірки гіпотез щодо отримання логіко-статистичних умов використання МНК.

Мірою адекватності моделі слугують відхилення фактичних значень від теоретичних $e_j = y_j - Y_j$. На величину цих відхилень впливає весь комплекс умов, зокрема:

- обсяг та однорідність сукупності;
- незалежність спостережень;
- інформативність включених у модель факторів;
- тип моделі.

Репрезентативність оцінок регресійного аналізу прямо пропорційна обсягу та однорідності сукупності. Саме недостатній обсяг сукупності та її неоднорідність вважаються найвагомішими чинниками неадекватності моделей. Тому при формуванні ознакової множини моделі слід враховувати співвідношення між обсягом вибірки і кількістю включених у модель факторів (воно має бути приблизно 8:1).

Оцінювання однорідності сукупності здійснюється на етапі розвідувального аналізу даних. Так, наявність аномальних значень, які не узгоджуються з розподілом основної маси даних, може бути наслідком помилок спостереження або результатом незвичайної комбінації причин і умов, у яких функціонує одиниця сукупності. Ідентифікація таких спостережень дає можливість усунути помилки, а якщо це неможливо, то вилучити аномальний об'єкт з подальшого аналізу. Якщо сукупність розшарована на групи (кластери), то в моделі можна врахувати таку неоднорідність.

Інформативність включених у модель факторних ознак залежить як від соціально-економічного змісту, так і від шкали вимірювання ознаки. Якщо ознака за змістом не інформативна, то ніякий спосіб моделювання не забезпечить належних результатів. Так само результати аналізу будуть суттєво різнитися залежно від того, якою шкалою представлено одну й ту саму ознаку (метричною, ранговою чи номінальною).

Ті властивості, що безпосередньо не вимірюються або не мають єдиного вимірника, включаються в модель у вигляді інтегральних оцінок. Наприклад, погодні умови характеризуються середньодобовою температурою повітря, кількістю опадів, тривалістю сонячного світла, хмарністю і т. ін. Усі ці характеристики агрегуються в індексі погодних умов.

Важливою умовою регресійного аналізу є відсутність мультиколінеарності, яка веде до зсунення оцінок параметрів моделі та унеможливорює коректну інтерпретацію результатів. Два фактори вважаються колінеарними, якщо коефіцієнт кореляції між ними перевищує сукупний коефіцієнт кореляції, тобто $r_{ik} > R$. Найпростіший спосіб усунення мультиколінеарності – виключити одну із корельованих ознак із моделі або замінити її іншою. Часом колінеарні фактори агрегуються в одну узагальнюючу оцінку.

Стабільність не включених у модель факторів означає, що вплив їх на варіацію у незначний і врівноважується, він однаковий в усіх частинах сукупності (умова гомоскедастичності). Математичною основою дотримання цих передумов МНК слугує ймовірнісний розподіл залишків e_j . Передбачається, що:

- для кожного спостереження залишок e_j – випадкова величина, яка має нормальний розподіл. Умова нормальності необхідна для визначення довірчих меж коефіцієнтів регресії і для перевірки гіпотез щодо їх істотності;
- математичне сподівання залишків $M(e) = 0$;
- дисперсія залишків однакова в усіх частинах сукупності: $s_e^2 = const$. Ця умова пов'язана з однорідністю сукупності;
- залишки незалежні, тобто відсутня серійна кореляція чи автокореляція даних.

2.2. Логістична регресія

2.2.1. Загальні поняття

Регресійні методи з'явилися як інтегральний компонент будь-якого аналізу даних, пов'язаного з описом взаємозв'язків між відповідними змінними та однією або більше незалежними змінними. Часто залежна змінна є дискретною, що набуває два або більше можливих значень. Останнім часом логістична регресія стала, в багатьох галузях, стандартним методом аналізу в таких ситуаціях.

Важливо зрозуміти, що мета аналізу, використовуючи даний метод, є такою ж самою, що й для будь-якої техніки, базованої на моделі: знайти найкраще наближення та найбільш змістовну модель, щоб описати взаємозв'язки між вихідними (залежними або відповідними) та множиною незалежних (предикативних або пояснюючих) змінних. Найбільш загальним методом моделювання є лінійна регресійна модель, де припускається, що вихідні змінні є неперервними.

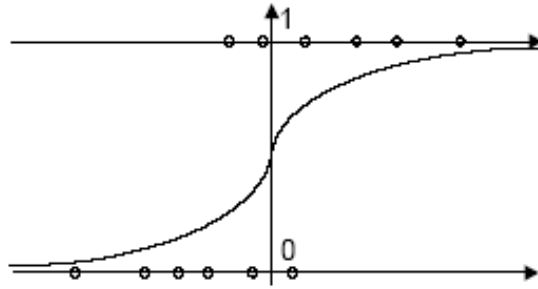
Основною відмінністю моделі логістичної регресії є те, що вихідна змінна в логістичній регресії є бінарною або дихотомічною. Ця відмінність між логістичною та лінійною регресією відбивається у виборі параметричної моделі та в припущеннях. Але методи, задіяні в аналізі, що використовує логістичну регресію, слідує тим самим основним принципам, що використовуються в лінійній регресії. Тому техніка, що використовується в лінійному регресійному аналізі, обумовлює підхід і до логістичної регресії. Можна відзначити дві основні відмінності. Перша стосується природи взаємозв'язків між вихідними та незалежними змінними. В будь-якій регресії ключовою величиною є середнє значення вихідної змінної. Ця величина називається умовним середнім та позначається як “ $E(Y|x)$ ”, де Y позначає вихідну змінну, а x – вектор значень незалежних змінних. Величина $E(Y|x)$ читається як „очікуване значення Y при даному значенні x ”. В лінійній регресії ми припускаємо, що це середнє значення може бути виражене як лінійне відносно x (або певного перетворення x або Y) рівняння, такого як:

$$E(X | x) = \beta_0 + \sum_{i=1}^p \beta_i x_i \quad (2.2.1.1)$$

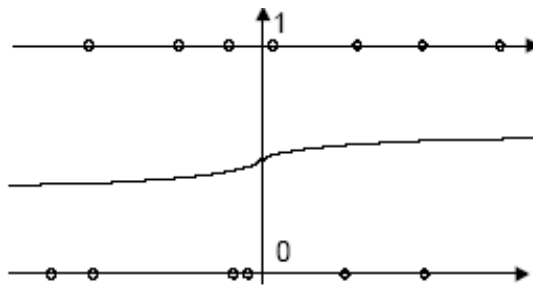
З цього виразу випливає, що $E(Y|x)$ може набувати будь-якого значення якщо x змінюється у проміжку від $-\infty$ до $+\infty$.

З дихотомічними даними умовне середнє має бути більшим або дорівнювати 0 та меншим або дорівнювати 1: $[0 \leq E(Y|x) \leq 1]$. Зміна $E(Y|x)$ на одиницю зміни x значно менша, коли умовне середнє близьке до 0 або 1. Така крива називається S-кривою. Вона нагадує графік функції розподілу випадкової величини. Тому не дивно, що використовуються певні відомі закони розподілу для побудови моделі для $E(Y|x)$ у випадку, коли Y є дихотомічною. В якості такого закону розподілу ми використаємо логістичний розподіл.

Мал. 2.2.1.1. Графік логістичної регресії з гарною класифікаційною здатністю



Мал. 2.2.1.2. Графік логістичної регресії з поганою класифікаційною здатністю



Існує дві основні причини для обрання логістичного розподілу. По-перше, з математичної точки зору, це достатньо гнучка функція, та, по-друге, вона дає легку для сприйняття інтерпретацію.

Щоб спростити запис, ми будемо використовувати позначення $\pi(x) = E(Y | x)$ для представлення умовного середнього Y при даному x , коли ми використовуємо логістичну регресію. Специфічна форма моделі логістичної регресії, яку ми будемо використовувати:

$$\pi(x) = \frac{e^{\beta_0 + \sum_{i=1}^p \beta_i x_i}}{1 + e^{\beta_0 + \sum_{i=1}^p \beta_i x_i}} \quad (2.2.1.2)$$

Перетворення $\pi(x)$, яке центральним у нашому дослідженні, називають логіт-перетворенням. Це перетворення визначається, у термінах $\pi(x)$, у вигляді:

$$g(x) = \ln \left[\frac{\pi(x)}{1 - \pi(x)} \right] = \beta_0 + \sum_{i=1}^p \beta_i x_i \quad (2.2.1.3)$$

Важливість цього перетворення полягає в тому, що $g(x)$ має багато бажаних властивостей моделі лінійної регресії. Логіт, $g(x)$, лінійний відносно параметрів, може бути неперервним, та може змінюватись у інтервалі від $-\infty$ до $+\infty$, в залежності від інтервалу, в якому лежить x .

Друга важлива відмінність між лінійною та логістичною регресією стосується умовного розподілу вихідної змінної. В лінійній регресії ми припускаємо, що спостереження вихідної змінної можуть бути виражені як $y = E(Y|x) + \varepsilon$. Величина ε називається похибкою та виражає спостережуване відхилення від умовного середнього. Найбільш загальним припущенням є те, що ε відповідає нормальному закону розподілу з нульовим середнім та певною дисперсією, що є сталою відносно рівнів незалежної змінної. З цього випливає, що умовний розподіл вихідної змінної при даному x буде нормальним із середнім $E(Y|x)$ та сталою дисперсією. У випадку з дихотомічною вихідною змінною ми можемо виразити значення вихідної змінної при даному x як $y = \pi(x) + \varepsilon$. Тут величина ε може приймати одне з двох можливих значень. Якщо $y=1$, тоді $\varepsilon = 1 - \pi(x)$ з ймовірністю $\pi(x)$, якщо $y=0$, то $\varepsilon = -\pi(x)$ з ймовірністю $1 - \pi(x)$. Тому ε має розподіл з середнім 0 та дисперсією, рівною $\pi(x)[1 - \pi(x)]$. Тому умовний розподіл вихідної змінної відповідає біноміальному розподілу з ймовірністю, заданою умовним середнім $\pi(x)$. [10]

Підсумовуючи, можна сказати, що в тому випадку, коли вихідна змінна є дихотомічною:

- Умовне середнє регресійного рівняння має бути обмеженим між 0 та 1. Ми зазначили, що модель логістичної регресії, $\pi(x)$ наведений у рівнянні (1.1) задовольняє обмеженню.
- Біноміальний, але не нормальний, розподіл описує розподіл похибок та буде статистичним розподілом, на якому базується аналіз.
- Принципами, якими керуються в аналізі, використовуючи лінійну регресію, керуються також в логістичній регресії.

2.2.2. Оцінювання параметрів логістичної регресії

Припустимо, ми маємо вибірку з n незалежних спостережень пар $(y_i; x_i)$, $i=1,2,\dots,n$, де y_i означає значення дихотомічної вихідної змінної та x_i - значення незалежної змінної для i -го спостереження. Більше того, припустимо, що вихідна змінна задана у вигляді 0 або 1, представляючи відсутність або присутність певної характеристики відповідно. Це кодування для дихотомічної змінної буде використовуватись у всьому тексті. Побудувати модель логістичної регресії відповідно рівняння (1.1) на множині даних ми маємо оцінити значення $\beta_0, \beta_1, \dots, \beta_p$ - невідомих параметрів.

В лінійній регресії для оцінювання невідомих параметрів найчастіше використовується метод найменших квадратів. В цьому методі ми вибираємо такі значення $\beta_0, \beta_1, \dots, \beta_p$, які мінімізують суму квадратів відхилення спостережуваних значень Y від прогнозованих значень, що базуються на моделі. Виходячи із звичайних припущень для лінійної регресії метод найменших квадратів надає оцінки з бажаними статистичними властивостями. На жаль, при застосуванні методу найменших квадратів з дихотомічними вихідними змінними оцінки не мають таких самих властивостей. [16]

Основним методом оцінювання, що призводить до функції з мінімальними квадратами у моделі лінійної регресії (коли похибки розподілені за нормальним законом) називається методом максимальної правдоподібності. Цей метод буде забезпечувати базис для нашого підходу до оцінювання моделі логістичної регресії. В загальному сенсі метод максимальної правдоподібності породжує значення невідомих параметрів, які максимізують ймовірність отримання спостережуваного набору даних. Щоб застосувати цей метод ми маємо сконструювати функцію, яка називається функцією правдоподібності. Ця функція виражає ймовірність спостережуваних даних як функцію невідомих параметрів. Оцінки максимальної правдоподібності цих параметрів вибираються таким чином,

щоб дані параметри максимізували цю функцію. Тому результуючі параметри є тими, що узгоджені найбільше із спостережуваними даними. Зараз ми опишемо, як отримати ці значення у моделі логістичної регресії. Якщо Y кодується як 0 та 1 то вираз для $\pi(x)$ наведений у рівнянні (2.1.1.2) забезпечує (для будь-яких значень $\beta = \beta_0, \beta_1, \dots, \beta_p$, вектор параметрів) умовну ймовірність, що $Y=1$ при даному x . Це може бути записано як $P(Y=1|x)$. З цього випливає, що величина $1 - \pi(x)$, що задає умовну ймовірність, що $Y=0$ при даному x , може бути записана як $P(Y=0|x)$. Тому, для тих пар $(x_i; y_i)$, де $y_i = 1$, внесок до функції правдоподібності, буде дорівнювати $\pi(x_i)$, та для тих пар, де $y_i = 0$, внесок до функції правдоподібності, буде дорівнювати $1 - \pi(x_i)$, де величина $\pi(x_i)$ означає значення функції $\pi(x)$, яке вона набуває на x_i . Зручним шляхом подання внеску до функції правдоподібності для пари $(x_i; y_i)$ є вираз:

$$\pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} \quad (2.2.2.1)$$

Оскільки припускається, що спостереження є незалежними, функція правдоподібності отримується як добуток компонент, наведених у рівнянні (2.1.2.1):

$$l(\beta) = \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} \quad (2.2.2.2)$$

Принцип максимальної правдоподібності покладає, що ми використовуємо як нашу оцінку β , значення, яке максимізує вираз (2.1.2.2). Але математично легше працювати з логарифмом виразу (2.1.2.2). Цей вираз, логарифм правдоподібності, визначається як:

$$L(\beta) = \ln[l(\beta)] = \sum_{i=1}^n \{y_i \ln[\pi(x_i)] + (1 - y_i) \ln[1 - \pi(x_i)]\} \quad (2.2.2.3)$$

Щоб знайти значення β , яке максимізує $L(\beta)$, ми диференціюємо $L(\beta)$ за β_0 та β_1 та покладаємо отримане значення рівне 0. Ці рівняння відомі як рівняння правдоподібності:

$$\sum_{i=1}^n [y_i - \pi(x_i)] = 0 \quad (2.2.2.4)$$

та

$$\sum_{i=1}^n x_{ij} [y_i - \pi(x_i)] = 0 \quad (2.2.2.5)$$

для $j=1,2,\dots,p$.

Знайдемо другу похідну:

$$\frac{\partial^2 L(\beta)}{\partial \beta_j^2} = - \sum_{i=1}^n x_{ij}^2 \pi_i (1 - \pi_i) \quad (2.2.2.7)$$

та

$$\frac{\partial^2 L(\beta)}{\partial \beta_j \partial \beta_l} = - \sum_{i=1}^n x_{ij} x_{il} \pi_i (1 - \pi_i) \quad (2.2.2.8)$$

для $j, l = 0, 1, 2, \dots, p$ де π_i означає $\pi_i(x)$. Нехай $(p+1) \times (p+1)$ матриця містить значення, інверсні відносно значень з виразів (2.3) та (2.4) та позначається як $I(\beta)$. Ця матриця називається інформаційною матрицею спостережень. Дисперсія та коваріації оцінюваних коефіцієнтів отримуються з оберненої матриці, яку ми позначимо $Var(\beta) = I^{-1}(\beta)$. За виключенням дуже специфічних випадків, неможливо записати точний вираз для елементів цієї матриці. Будемо використовувати запис $Var(\beta_j)$, щоб позначити j -й діагональний елемент матриці, що є дисперсією $\hat{\beta}_j$, та $Cov(\beta_j, \beta_l)$, щоб позначити довільний недіагональний елемент, що є коваріацією $\hat{\beta}_j$ та $\hat{\beta}_l$. Оцінки дисперсії та коваріації, які можуть бути записані як $Var(\hat{\beta})$, отримуються з розрахунку $Var(\hat{\beta})$ при $\hat{\beta}$. [17] Ми будемо використовувати $Var(\hat{\beta}_j)$ та $Cov(\hat{\beta}_j, \hat{\beta}_l)$, $j, l = 0, 1, 2, \dots, p$, щоб позначити значення цієї матриці. Для більшої частини ми будемо мати можливість використовувати тільки оцінені стандартні похибки оцінюваних коефіцієнтів, які можна записати як:

$$SE(\hat{\beta}_j) = [Var(\hat{\beta}_j)]^{1/2} \quad (2.2.2.9)$$

для $j=0,1,2,\dots,p$.

Формулювання інформаційної матриці, яке буде корисним, коли ми будемо аналізувати побудову моделі та оцінювання якості є $\hat{I}(\hat{\beta}) = X'VX$, де X матриця $n \times (p+1)$, та V - матриця $n \times n$ з діагональними елементами $\hat{\pi}_i \times (1 - \hat{\pi}_i)$. Матриця X :

$$X = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix}$$

та матриця V :

$$V = \begin{bmatrix} \hat{\pi}_1(1 - \hat{\pi}_1) & 0 & \dots & 0 \\ 0 & \hat{\pi}_2(1 - \hat{\pi}_2) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \hat{\pi}_n(1 - \hat{\pi}_n) \end{bmatrix}$$

Тоді розв'язок системи (2.1.2.4), (2.1.2.5) знаходиться чисельно, наприклад, методом Ньютона. Рекурсивне рівняння для знаходження параметрів:

$$\beta^{(t+1)} = \beta^t + (X'VX)^{-1} X'(y - \pi^t) \quad (2.2.2.10)$$

Оскільки цільова функція опукла, то метод завжди збігається. [11]

2.2.3 Оцінка адекватності моделі

Для аналізу якості моделі, крім значення функції правдоподібності, можна використати наступні критерії:

- Інформаційний критерій Акайке:

$$AIC = -\frac{2}{N} (L(\hat{\beta}) - p) \quad (2.2.3.1)$$

- Для моделей із залежною бінарною змінною – аналог коефіцієнта детермінації – *псевдо* – R^2

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2 + N\sigma^2}, \quad (2.2.3.2)$$

де $\sigma^2 = \frac{\pi^2}{3}$ - дисперсія похибки ε .

Проведемо аналіз значимості змінних моделі. Значимість перевіряється шляхом порівняння моделі, що враховує певну змінну, та моделі, що не враховує данну змінну. В логістичній регресії порівняння спостережних та прогнозованих значень базується на логарифмі функції правдоподібності. Вводиться так звана насичена модель (тобто така модель, яка має таку ж саму кількість параметрів, як кількість спостережень). Порівняння спостережних та прогнозних значень, використовуючи функцію правдоподібності, базується на наступному виразі:

$$D = -2 \ln \left[\frac{\text{правдоподібність} \text{ _ оцінюваної _ моделі }}{\text{правдоподібність} \text{ _ насиченої _ моделі }} \right] \quad (2.2.3.3)$$

Величина в дужках називається показником правдоподібності. Використовуючи мінус подвійний логарифм, щоб отримати величину, розподіл якої відомий, і може бути використаний в цілях перевірки статистичних гіпотез. Такий тест має назву тест показника правдоподібності. Використовуючи вираз (2.1.3.3) та (2.1.2.3):

$$D = -2 \sum_{i=1}^n \left[y_i \ln \left(\frac{\hat{\pi}_i}{y_i} \right) + (1 - y_i) \ln \left(\frac{1 - \hat{\pi}_i}{1 - y_i} \right) \right] \quad (2.2.3.3)$$

де $\hat{\pi}_i = \hat{\pi}(x_i)$

Статистику D інколи називають девіацією. Ця статистика відіграє провідну роль в деяких підходах до оцінювання ступеня узгодженості моделі. Девіація для логістичної регресії грає ту ж саму роль, що сума квадратів залишків у лінійній регресії. [18]

У випадку, коли вихідна змінна приймає значення 0 або 1, правдоподібність насиченої моделі дорівнює 1. Це впливає з визначення насиченої моделі що $\hat{\pi}_i = y_i$ та правдоподібність:

$$l(\text{насиченої} \text{ _ моделі}) = \prod_{i=1}^n y_i^{y_i} \times (1 - y_i)^{(1-y_i)} = 1 \quad (2.2.3.4)$$

Тому з рівняння (2.1.3.3) випливає, що девіація:

$$D = -2\ln(\text{правдоподібність оцінюваної моделі}) \quad (2.2.3.5)$$

З метою оцінювання значимості незалежної змінної ми порівняємо значення D з та без незалежної змінної у рівнянні. Зміна D , пов'язана з включенням незалежної змінної:

$$G = D(\text{модель без змінної}) - D(\text{модель зі змінною}) \quad (2.2.3.6)$$

Ця статистика грає ту ж саму роль в логістичній регресії, що числівник часткової F у лінійній регресії. Оскільки правдоподібність насиченої моделі є однаковою для обох значень D , ми можемо записати:

$$G = -2\ln\left[\frac{\text{правдоподібність без змінної}}{\text{правдоподібність із змінною}}\right] \quad (2.2.3.7)$$

Надалі ми висуваємо гіпотезу, що розглядувана змінна рівна нулю (нуль-гіпотеза). Статистика G розподілена за χ^2 -розподілом з $p+1$ ступенями свободи. Тому ми перевіряємо значимість нуль-гіпотези шляхом порівняння теоретичного та емпіричного значень χ^2 -розподілу.

Іншим шляхом визначення значимості змінної є побудова статистика Вальда:

$$W_j = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \quad (2.2.3.8)$$

Провівши перетворення, отримаємо:

$$W = \hat{\beta}'[Var(\hat{\beta})]^{-1}\hat{\beta} = \hat{\beta}'(X'VX)\hat{\beta} \quad (2.2.3.9)$$

Дана статистика розподілена за χ^2 -розподілом з $p+1$ ступенями свободи. Для кожної змінної також перевіряється нуль-гіпотеза. [12]

2.2.4. Критерії Шварца і Акайке

При побудові моделі, що адекватно описує досліджуваний процес в економіці, дуже важливу роль грає аналіз правильності її специфікації. Негативно на пояснюючих властивостях моделі позначається як відсутність значимої змінної, так і надлишкова присутність незначущої.

У випадку, коли в модель не включена істотна змінна (істотною називають змінну, котра повинна бути в моделі відповідно до правильної теорії), спостерігаються наступні наслідки:

1. Зникає можливість правильної оцінки й інтерпретації рівнянь.
2. Коефіцієнти при змінних, що залишилися, стають зміщеними.
3. Стандартні помилки коефіцієнтів і t-статистики некоректні й не можуть бути використані для судження про якість підгонки пропонованої моделі.

Наприклад, припустимо, що з моделі:

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i \quad (2.2.4.1)$$

виключена змінна X_2 . Тоді в новій специфікації фактично розглядається модель:

$$Y_i = \alpha + \beta_1 X_{1i} + u_i \quad (2.2.4.2)$$

$$\text{де } u_i = \beta_2 X_{2i} + \varepsilon_i. \quad (2.2.4.3)$$

Якщо пояснюючі змінні X_1 і X_2 корельовані, то порушується передумова теореми Гаусса-Маркова про некорельованість випадкового члена і регресорів, оскільки в цьому випадку між X_1 і u існує ненульова кореляція. Оцінки, отримані по методу найменших квадратів для даної моделі, уже не є ефективними серед лінійних оцінок.

Оцінки навіть не є незміщеними, оскільки для МНК оцінки коефіцієнта у цьому випадку:

$$\hat{\beta}_1 = \beta_1 + \beta_2 \frac{\text{Cov}(X_1, X_2)}{\text{Var}(X_1)}. \quad (2.2.4.4)$$

Спостерігається зсув $\beta_2 \frac{\text{Cov}(X_1, X_2)}{\text{Var}(X_1)}$.

Включення несуттєвої змінної в модель не приводить до зсуву оцінок коефіцієнтів, але з'являється інший недолік - ростуть стандартні помилки коефіцієнтів. Оцінки стають статистично незначущими.

Якщо точна специфікація моделі невідома (що практично завжди й буває), то користуються критеріями, що дозволяють вибирати з деякої множини моделей найкращу. [13]

Найпоширенішими критеріями є критерій Шварца (Schwarz) і критерій Акайке (Akaike). Обидва критерії дозволяють вибирати найкращу модель із множини різних специфікацій. Критерії чисельно побудовані так, щоб урахувати вплив на якість підгонки моделі двох протилежних тенденцій.

При додаванні змінних у модель якість підгонки в загальному випадку збільшується. Помітимо, що число регресорів повинне бути розумним, щоб не викликати "штучної підгонки" залежної змінної пояснюючими. З іншого боку, недостатнє включення змінних у модель дає більшу стандартну помилку, і якість підгонки знижується.

Формули для розрахунку критеріїв Akaike і Schwarz:

$$Akaike\ Information\ Criterion = \ln(\sigma^2) + \frac{2K}{n}, \quad (2.2.4.5)$$

$$Schwarz\ Criterion = \ln(\sigma^2) + \frac{K \ln n}{n}, \quad (2.2.4.6)$$

де $\sigma^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n}$ - вибіркова дисперсія, DO – число обмежень на ступені свободи.

Значення K у цьому випадку дорівнює числу незалежних змінних, включаючи вільний член. Таким чином, якщо в моделі присутня два регресора та вільний член, то число обмежень на ступені свободи буде дорівнювати трьом.

Перший доданок являє собою штраф за більшу дисперсію, другий - штраф за використання додаткових змінних. Критерії розраховуються для кожної розглянутої специфікації. При порівнянні двох типів моделей перевага віддається специфікації, що має найменші значення критеріїв.

2.2.5 Побудова порогу відсікання та ROC-аналіз

ROC-Крива (Receiver Operator Characteristic) – крива, що найбільше часто використовується для подання результатів бінарної класифікації в машинному навчанні. Назва прийшла із систем обробки сигналів. Оскільки класів два, один з них називається класом з позитивними результатами, другий – з негативними результатами. ROC-Крива показує залежність кількості вірно класифікованих позитивних прикладів від кількості невірно класифікованих негативних прикладів. У термінології ROC-Аналізу перші називаються істинно позитивним, другі – неправдиво негативною множиною. При цьому передбачається, що в класифікатора є деякий параметр, варіюючи який, ми будемо одержувати ту або іншу розбивку на два класи. Цей параметр часто називають порогом відсікання. Залежно від його значень будуть отримані різні величини помилок I і II роду. [14]

У логістичній регресії поріг відсікання змінюється від 0 до 1 - це і є розрахункове значення рівняння регресії. Будемо називати його рейтингом. Для розуміння суті помилок I і II роду розглянемо чотирипільну таблицю спряженості, що будується на основі результатів класифікації моделлю й фактичної (об'єктивної) приналежності прикладів до класів.

Табл. 2.2.5.1. Класифікація випадків прогнозування

	Фактично	
Модель	позитивно	негативно
позитивно	TP	FP
негативно	FN	TN

- TP (*True Positives*) – вірно класифіковані позитивні приклади (так звані істинно позитивні випадки);
- TN (*True Negatives*) – вірно класифіковані негативні приклади (істинно негативні випадки);
- FN (*False Negatives*) – позитивні приклади, класифіковані як негативні (помилка I роду). Це так званий "помилковий пропуск" - коли цікавляча нас подія помилково не виявляється (неправдиво негативні приклади);
- FP (*False Positives*) – негативні приклади, класифіковані як позитивні (помилка II роду); Це помилкове виявлення, тому що при

відсутності події помилково виноситься рішення про його присутність (неправдиво позитивні випадки).

Що є позитивною подією, а що - негативною, залежить від конкретної задачі.

При аналізі частіше оперують не абсолютними показниками, а відносними - частками, вираженими у відсотках:

Частка істинно позитивних прикладів:

$$TPR = \frac{TP}{TP + FN} \quad (2.2.5.1)$$

Частка неправдиво позитивних прикладів:

$$FPR = \frac{FP}{TN + FP} \quad (2.2.5.2)$$

Введемо ще два визначення: чутливість і специфічність моделі. Ними визначається об'єктивна цінність будь-якого бінарного класифікатора.

Чутливість - це і є частка істинно позитивних випадків:

$$Se = \frac{TP}{TP + FN} \quad (2.2.5.3)$$

Специфічність - частка істинно негативних випадків, які були правильно ідентифіковані моделлю:

$$Sp = \frac{TN}{TN + FP} \quad (2.2.5.4)$$

Відзначимо, що $FPR = 1 - Sp$.

Модель із високою чутливістю часто дає істинний результат при наявності позитивного результату (виявляє позитивні приклади). Навпаки, модель із високою специфічністю частіше дає істинний результат при наявності негативного результату (виявляє негативні приклади).

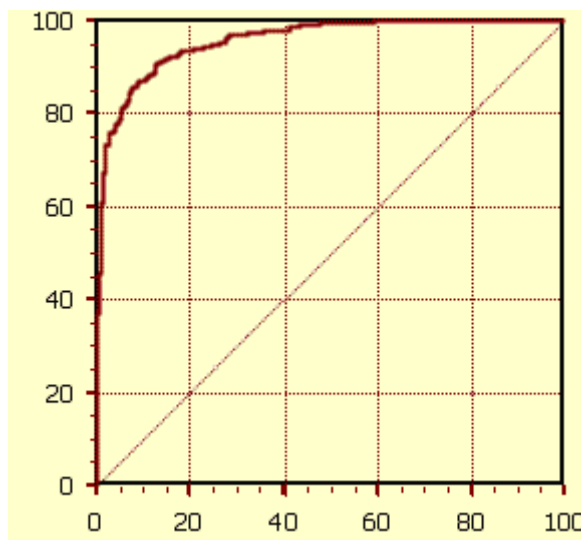
ROC-крива будується у такий спосіб:

1. Для кожного значення порогу відсікання, що міняється від 0 до 1 із кроком dx (наприклад, 0.01) розраховуються значення чутливості Se і специфічності Sp . Як альтернатива порогом може бути кожне наступне значення приклада у вибірці.

2. Будується графік залежності: по осі Y відкладається чутливість Se , по осі X - $1-Sp$ (сто відсотків мінус специфічність), або, що те ж саме, FPR - частка неправдиво позитивних випадків.

ROC-крива має наступний вигляд:

Мал. 2.2.5.1 ROC-крива



Для ідеального класифікатора графік ROC-кривої проходить через верхній лівий кут, де частка істинно позитивних випадків становить 100% або 1.0 (ідеальна чутливість), а частка неправдиво позитивних прикладів дорівнює нулю. Тому чим ближче крива до верхнього лівого кута, тим вище передбачувальна здатність моделі. Навпаки, чим менше вигин кривої й чим ближче вона розташована до діагональної прямій, тим менш ефективна модель. Діагональна лінія відповідає "марному" класифікатору, тобто повної нерозрізненості двох класів.

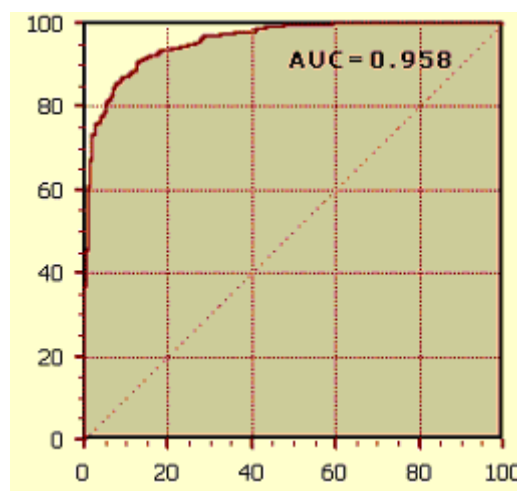
При візуальній оцінці ROC-кривих розташування їх одна відносно іншої вказує на їхню порівняльну ефективність. Крива, розташована вище та лівіше свідчить про більшу передбачувальну здатність моделі.

Візуальне порівняння кривих ROC не завжди дозволяє виявити найбільш ефективну модель. Своєрідним методом порівняння ROC-Кривих є оцінка площі під кривими. Теоретично вона змінюється від 0 до 1.0, але, оскільки модель завжди характеризується кривій, розташованої вище позитивної діагоналі, то звичайно говорять про зміни від 0.5 ("марний" класифікатор) до 1.0 ("ідеальна" модель). Ця оцінка може бути отримана безпосередньо обчисленням площі під багатогранником, обмеженим праворуч і знизу

осями координат і ліворуч угорі - експериментально отриманими точками. Чисельний показник площі під кривою називається AUC (Area Under Curve). Обчислити його можна, наприклад, за допомогою чисельного методу трапецій:

$$AUC = \int f(x)dx = \sum_i \left[\frac{X_{i+1} + X_i}{2} \right] \cdot (Y_{i+1} - Y_i) \quad (2.2.5.5)$$

Мал. 2.2.5.2. Площа під ROC-кривою



З деякими припущеннями можна вважати, що чим більше показник AUC, тим кращою прогностичною силою володіє модель. Однак варто зазначити, що:

- показник AUC призначений скоріше для порівняльного аналізу декількох моделей;
- AUC не містить ніякої інформації про чутливість і специфічність моделі.

У літературі іноді приводиться наступна експертна шкала для значень AUC, по якій можна судити про якість моделі:

Табл. 2.2.5.2. Емпіричне визначення якості моделі

Інтервал AUC	Якість моделі
0.9-1.0	Відмінна
0.8-0.9	Дуже добра

0.7-0.8	Добра
0.6-0.7	Середня
0.5-0.6	Незадовільна

Ідеальна модель володіє 100% чутливістю й специфічністю. Однак на практиці домогтися цього неможливо, більше того, неможливо одночасно підвищити і чутливість, і специфічність моделі. Компроміс знаходиться за допомогою порогу відсікання, тому що граничне значення впливає на співвідношення Se і Sp . Можна говорити про задачу знаходження оптимального порогу відсікання. [19]

Поріг відсікання потрібний для того, щоб застосовувати модель на практиці: відносити нові приклади до одному із двох класів. Для визначення оптимального порогу потрібно задати критерій його визначення, тому що в різних задачах присутня своя оптимальна стратегія.

Критеріями вибору порогу відсікання можуть виступати:

- Вимога мінімальної величини чутливості (специфічності) моделі. Наприклад, потрібно забезпечити чутливість тесту не менш 80%. У цьому випадку оптимальним порогом буде максимальна специфічність (чутливість), що досягається при 80% (або значення, близьке до нього "праворуч" через дискретність ряду) чутливості (специфічності).
- Вимога максимальної сумарної чутливості й специфічності моделі, тобто:

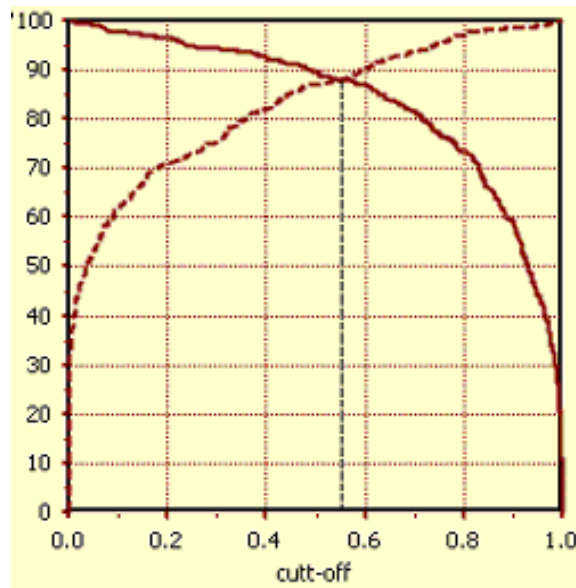
$$Cut_off_0 = \arg \max_k (Se_k + Sp_k) \quad (2.2.5.6)$$

- Вимога балансу між чутливістю й специфічністю, тобто коли $Se \approx Sp$:

$$Cut_off_0 = \arg \min_k |Se_k - Sp_k| \quad (2.2.5.7)$$

Другий варіант порогу звичайно пропонується користувачеві за замовчуванням. У третьому випадку поріг є крапка перетинання двох кривих, коли по осі X відкладається поріг відсікання, а по осі Y -чутливість або специфічність моделі:

Мал. 2.2.5.3. Оптимальний баланс між специфічністю та чутливістю



Надалі, при проведенні експериментів, будемо користуватись саме критерієм (2.2.5.6).

Існують і інші підходи, коли помилкам I і II роду призначається вага, що інтерпретується як ціна помилок. Але тут постає проблема визначення цієї ваги, що саме по собі є складною, а часто не розв'язною задачею. [15]

2.2.6 Застосування логістичної регресії

Розглянемо аспекти практичного застосування логістичної регресії для оцінки кредитоспроможності позичальника. Початковими даними для побудови логістичної регресії є кредитна історія портфелю кредитів – пари (X_i, Y_i) , $i=1, 2, \dots, m$.

Можемо побудувати логістичну регресію:

$$Y = \frac{e^{\beta_0 + \sum_{i=1}^l \beta_i x_i}}{1 + e^{\beta_0 + \sum_{i=1}^l \beta_i x_i}} \quad (2.2.6.2)$$

Звідки знаходимо параметри β_0, \dots, β_l та поріг відсікання Cut_off , за допомогою максимізації критерію (2.1.5.6):

$$Cut_off = \arg \max_k (Se_k + Sp_k) \quad (2.2.6.3)$$

Даний рейтинг Y порівнюємо з Cut_off . Якщо рейтинг не перевищує поріг відсікання, то кредит надається, якщо перевищує – клієнту відмовляють.

2.3. Розробка моделі

2.3.1. Вхідні параметри моделі

В даній роботі пропонується модель для оцінювання кредитоспроможності корпоративних клієнтів банку. Хоча й у європейській практиці немає широкої практики використання скорингових моделей для оцінки корпоративних клієнтів (такі клієнти потребують більш детального розгляду і часто остаточна оцінка надається кредитними експертами), проте ці моделі спроможні описати загальне становище справ клієнта і надати попередню оцінку для подальших висновків (наприклад, кредитного комітету). В якості вхідних параметрів було взято фінансові коефіцієнти, що в сукупності можуть дати чітку картину становища справ позичальника, кредитна заява якого перебуває на розгляді або такого, що піддається переоцінці. [20]

В пропонуємій моделі в якості вхідних будемо використовувати наступні параметри:

- коефіцієнт загальної ліквідності;
- коефіцієнт поточної ліквідності;
- коефіцієнт миттєвої ліквідності;
- коефіцієнт незалежності;
- коефіцієнт ділової активності;
- коефіцієнт маневреності власних коштів;
- коефіцієнт рентабельності активів;
- коефіцієнт рентабельності продаж;
- період обіговості оборотних активів;
- коефіцієнт достатності робочого капіталу;

- коефіцієнт покриття зобов'язань EBITDA.

Частина з них описана в Розділі 1.4 Оцінка кредитоспроможності юридичних осіб. Дано роз'яснення щодо решти коефіцієнтів.

Коефіцієнт ділової активності показує відношення виручки до валюти балансу:

$$KDA = \frac{Bp}{Bб} \quad (2.3.1.1)$$

де Bp – виручка позичальника,

$Bб$ – валюта балансу.

Період обіговості оборотних активів відображає термін, за який оборотні активи проходять повний оборотний цикл:

$$ПОБ = \frac{Зп}{Сб} \quad (2.3.1.2)$$

де $Зп$ – запаси позичальника,

$Сб$ – собівартість його продукції.

Коефіцієнт достатності робочого капіталу показує частку капіталу позичальника до його необоротних активів:

$$KDP = \frac{(Bк + Дз)}{Ha} \quad (2.3.1.3)$$

де $Bк$ – власний капітал позичальника,

$Дз$ – довгострокові зобов'язання,

Ha – необоротні активи.

Коефіцієнт покриття зобов'язань EBITDA показує відношення значення EBITDA до зобов'язань позичальника:

$$KПЗЕ = \frac{EBITDA}{Зб} \quad (2.3.1.5)$$

де $Зб$ – зобов'язання позичальника.

Особливості вищевказаних коефіцієнтів полягають в тому, що не завжди вдається зазначити чіткі межі, при яких можна констатувати, що діяльність позичальника є цілком прийнятною і банк може надати йому кредит. Крім

того кредитоспроможність не завжди лінійно залежить від значення деяких коефіцієнтів (тобто при збільшення або зменшенні коефіцієнта кредитоспроможність не завжди змінюється монотонно). Тому було прийнято рішення брати в якості вхідних параметрів не самі значення коефіцієнтів, а їх бальні оцінки. Пропонується як бали використовувати значення ймовірності дефолту позичальника.

2.3.2. Бальна оцінка коефіцієнтів

Постає питання як найкраще оцінити вагу того чи іншого коефіцієнта. Як рішення даної проблеми пропонується розбивати інтервал можливих значень параметрів на відрізки із відповідними балами (ймовірності дефолту), які ставляться у відповідність коефіцієнтам та використовуватимуться в якості вхідних параметрів моделі. [21]

У нас є набір величин (X_i, Y_i) , $i=1, 2, \dots, m$, де $X = (x_1, \dots, x_n)$ - вектор параметрів позичальника, Y - змінна, що характеризує кредитоспроможність позичальника або його дефолт (приймає значення на відрізу $[0;1]$, 0 – діяльність позичальника відмінна, 1 – позичальник не кредитоспроможний). Виберемо 1-й параметр позичальника – коефіцієнт загальної ліквідності X_1 і відповідні значення ймовірності дефолту Y . Використовуючи дані кредитного портфелю ставимо у відповідність значенню коефіцієнта кожного позичальника значення ймовірності дефолту (0 – якщо позичальник виконує вимоги кредитної угоди, 1 – має прострочку і не в змозі виконувати свої зобов'язання перед банком). Отже, маємо набір даних (X_1, Y) . Відсортуюмо цю вибірку за зростанням X_1 і відкинемо граничні дані, які можуть негативно вплинути на передбачувальну здатність моделі (окремі надто великі або малі значення порушують загальний розподіл значень коефіцієнтів). Спробуємо розбити можливі значення коефіцієнта X_1 на інтервали, відслідковуючи залежність між значеннями коефіцієнта і ймовірністю дефолту.

Встановимо вимоги до шуканих інтервалів:

- 1) інтервал має містити не менше 5% вибірки;
- 2) інтервал має бути не вужче 0,1.

Шукаємо точку інтервалу, в якій ймовірність дефолту приймає найменше значення, поступово збільшуючи інтервал шляхом додавання наступного елемента вибірки. Ймовірність дефолту розраховуємо наступним чином:

$$pd = \frac{\sum_{i=1}^k Y_i}{k} \quad (2.3.2.1)$$

де Y_i – ймовірності дефолтів відповідних позичальників,

k – кількість позичальників, взятих до розгляду.

Аналогічно знаходимо точку, в якій ймовірність дефолту приймає максимальне значення. Маємо:

$$X_{\min} = \arg \min_{a < X < b} pd \quad (2.3.2.2)$$

$$X_{\max} = \arg \max_{a < X < b} pd \quad (2.3.2.3)$$

де $[a; b]$ – відрізок можливих значень коефіцієнта X_1 .

З отриманих двох точок вибираємо меншу x_1 – таким чином отримали перший інтервал $[a; x_1]$.

$$x_1 = \min(X_{\min}, X_{\max}) \quad (2.3.2.3)$$

Повторюємо ці дії до тих пір, поки інтервали не накриють усю вибірку.

Тобто:

$$X_{\min}^1 = \arg \min_{x_1 < X < b} pd \quad (2.3.2.4)$$

$$X_{\max}^1 = \arg \max_{x_1 < X < b} pd \quad (2.3.2.5)$$

$$x_2 = \min(X_{\min}^1, X_{\max}^1) \quad (2.3.2.6)$$

.....

$$X_{\min}^{n-1} = \arg \min_{x_{n-1} < X < b} pd \quad (2.3.2.7)$$

$$X^{n-1}_{\max} = \arg \max_{q_{n-1} < X < b} pd \quad (2.3.2.8)$$

$$x_n = \min(X^{n-1}_{\min}, X^{n-1}_{\max}) \quad (2.3.2.9)$$

Отже, отримали набір інтервалів: $[a; x_1], (x_1; x_2], \dots, (x_n; b]$. Оскільки a і b – граничні можливі значення коефіцієнта X_1 , то для подальшого аналізу можемо використовувати наступні інтервали: $(-\infty; x_1], (x_1; x_2], \dots, (x_n; +\infty)$.

Тепер маємо змогу замість значень коефіцієнта X_1 підставити відповідні значення ймовірностей дефолту згідно з тим, в який саме інтервал потрапляє значення коефіцієнта того чи іншого позичальника. Таким чином отримали нову вибірку (pd_{X_1}, Y) .

Аналогічним чином отримаємо значення ймовірностей дефолту для інших коефіцієнтів: $pd_{X_2}, \dots, pd_{X_{11}}$. В результаті маємо вибірку на основі якої матимемо можливість побудувати модель, а саме:

$$(pd_i^{X_k}, Y_i), i=1, \dots, m; k=1, \dots, 11. \quad (2.3.2.10)$$

РОЗДІЛ 3

РЕАЛІЗАЦІЯ МОДЕЛІ

3.1. Опис вхідних даних

Для аналізу кредитоспроможності позичальників було вибрано наступні параметри:

- код ЄДРПОУ позичальника;
- ймовірність дефолту позичальника;
- коефіцієнт загальної ліквідності;
- коефіцієнт поточної ліквідності;
- коефіцієнт миттєвої ліквідності;
- коефіцієнт незалежності;
- коефіцієнт ділової активності;
- коефіцієнт маневреності власних коштів;
- коефіцієнт рентабельності активів;
- коефіцієнт рентабельності продаж;
- період обіговості оборотних активів;
- коефіцієнт достатності робочого капіталу;
- коефіцієнт покриття зобов'язань EBITDA.

Експерименти проводились на наступній вибірці:

- реальна вибірка одного з комерційних банків України.
Кількість позичальників:
 - тих, що повернули кредит – 470;
 - тих, що не повернули кредит – 92.

Було проведено 2 експерименти з метою оцінки впливу попередньої обробки даних шляхом розбиття на інтервали з відповідними бальними оцінками (ймовірності дефолту) на якість моделі. Для цього спочатку було побудовано модель на основі реальних значень фінансових коефіцієнтів

позичальника, а потім – модель на основі бальних оцінок тих же коефіцієнтів. Для побудови моделі було застосовано логістичну регресію з використанням пакету EViews 3.1.

3.2. Попередня обробка даних

Алгоритм попередньої обробки даних (розбиття на інтервали з бальними оцінками) наведений у 2.3.2. Застосуємо його до нашої експериментальної вибірки. Для цього було написано програму на мові Visual Basic 6.0 з досить простим і зручним інтерфейсом. Отримаємо:

Мал. 3.2.1. Коефіцієнти ліквідності

Коефіцієнт загальної ліквідності K1			Коефіцієнт поточної ліквідності K2			Коефіцієнт миттєвої ліквідності K3		
ліва грань	права грань	PD	ліва грань	права грань	PD	ліва грань	права грань	PD
0.046677	0.565966	0.488	0.004698	0.16338	0.647	0	0.038686	0.237
0.565966	0.811477	0.091	0.16338	0.257861	0.355	0.038686	0.080357	0.095
0.811477	0.991223	0.127	0.257861	0.369897	0.083	0.080357	0.208815	0.011
0.991223	1.140898	0.225	0.369897	0.481109	0.244	0.208815	0.346299	0.138
1.140898	1.27458	0.154	0.481109	0.575065	0.143	0.346299	7.154313	0.026
1.27458	1.448688	0.083	0.575065	0.672555	0.056			
1.448688	1.577442	0.207	0.672555	0.7722	0.186			
1.577442	1.876811	0.022	0.7722	0.886609	0.074			
1.876811	2.126282	0.250	0.886609	1.000465	0.227			
2.126282	2.630571	0.125	1.000465	1.149902	0.026			
2.630571	3.575576	0.114	1.149902	1.408503	0.207			
3.575576	9.957313	0.000	1.408503	1.955875	0.026			
			1.955875	7.804237	0.059			

Мал. 3.2.2. Коефіцієнти незалежності і ділової активності

Коефіцієнт незалежності K4			Коефіцієнт ділової активності K5			Коефіцієнт маневреності власних коштів K6		
ліва грань	права грань	PD	ліва грань	права грань	PD	ліва грань	права грань	PD
0.059569	4.26257	0.199	0.0094	0.551173	0.341	-76.1889	-7.79019	0.097
4.26257	8.291225	0.133	0.551173	2.048034	0.156	-7.79019	-3.59007	0.182
8.291225	13.86032	0.051	2.048034	2.577254	0.139	-3.59007	-1.94312	0.033
13.86032	24.99326	0.125	2.577254	3.119512	0.118	-1.94312	-0.87525	0.098
24.99326	548.8199	0.158	3.119512	3.829858	0.045	-0.87525	0.000384	0.225
			3.829858	4.636856	0.135	0.000384	0.998752	0.168
			4.636856	5.634486	0.027			
			5.634486	7.493498	0.140			
			7.493498	11.56278	0.069			

11.56278 64.04109 0.063

Мал. 3.2.3. Коефіцієнти рентабельності і обіговості

Коефіцієнт рентабельності активів K7			Коефіцієнт рентабельності продаж K8			Період обіговості оборотних активів K9		
ліва грань	права грань	PD	ліва грань	права грань	PD	ліва грань	права грань	PD
-0.31458	-0.06279	0.548	-9.50609	-0.11765	0.500	5.303997	140.489	0.098
-0.06279	-0.00715	0.235	-0.11765	-0.00716	0.310	140.489	280.6613	0.187
-0.00715	0.006028	0.183	-0.00716	0.070393	0.119	280.6613	461.3728	0.245
0.006028	0.019196	0.167	0.070393	0.133176	0.310	461.3728	646.7257	0.621
0.019196	0.033266	0.097	0.133176	4.6668	0.123	646.7257	16493.64	0.180
0.033266	0.048227	0.163						
0.048227	0.062519	0.029						
0.062519	0.084243	0.200						
0.084243	0.117389	0.065						
0.117389	0.165602	0.138						
0.165602	0.992417	0.081						

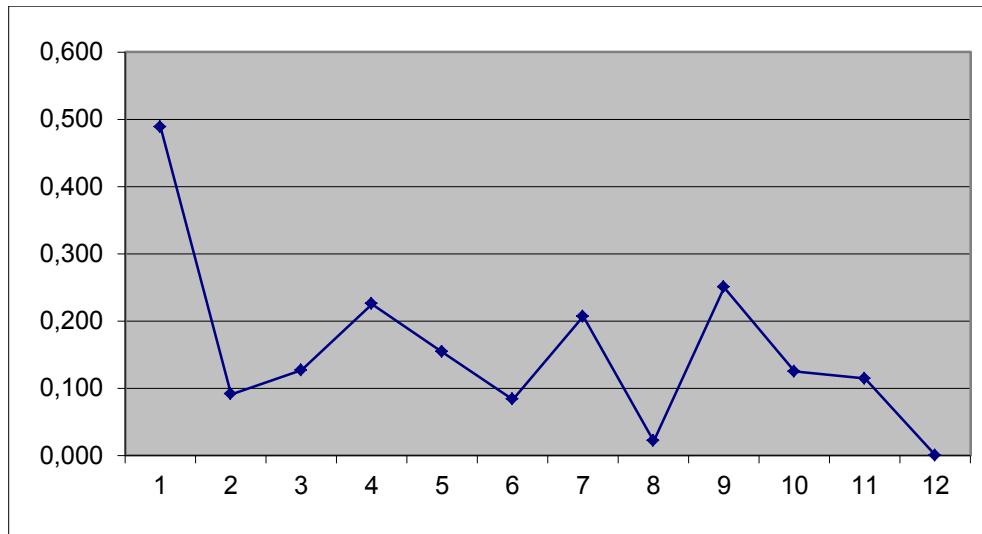
Мал. 3.2.4. Коефіцієнти достатності капіталу і покриття зобов'язань

Коефіцієнт достатності робочого капіталу K10			Коефіцієнт покриття зобов'язань EBITDA K11		
ліва грань	права грань	PD	ліва грань	права грань	PD
0.012955	10.08724	0.162	0	0.06337	0.198
10.08724	25.58296	0.276	0.06337	0.142701	0.126
25.58296	801.3178	0.105	0.142701	0.244028	0.192
			0.244028	0.317631	0.060
			0.317631	0.399005	0.267
			0.399005	0.611046	0.154
			0.611046	1.08594	0.323
			1.08594	7.227599	0.063

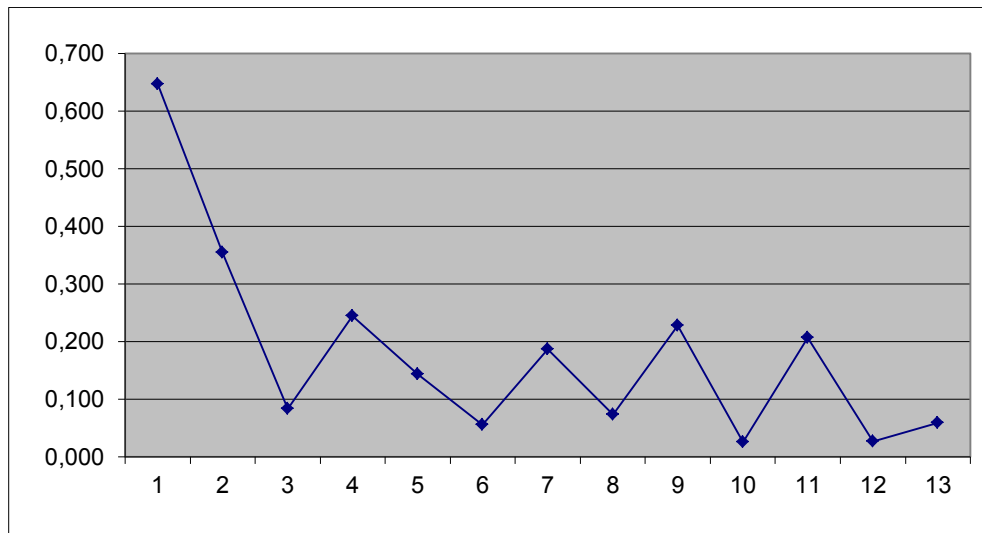
Отримали інтервали для фінансових коефіцієнтів з відповідними значеннями ймовірності дефолту, які будемо застосовувати в якості їх бальних оцінок. На основі їх значень маємо можливість побудувати модель для оцінки кредитоспроможності позичальників банку. Для цього замінимо реальні значення фінансових коефіцієнтів на отримані значення ймовірностей дефолту залежно від потрапляння коефіцієнту в той чи інший інтервал.

Спробуємо проаналізувати поведінку кредитоспроможності позичальника в залежності від значень отриманих оцінок для кожного коефіцієнта. Для наочності побудуємо графіки цих оцінок. На основі поведінки графіку можемо визначити характер впливу окремого фінансового коефіцієнта на кредитоспроможність позичальника в цілому.

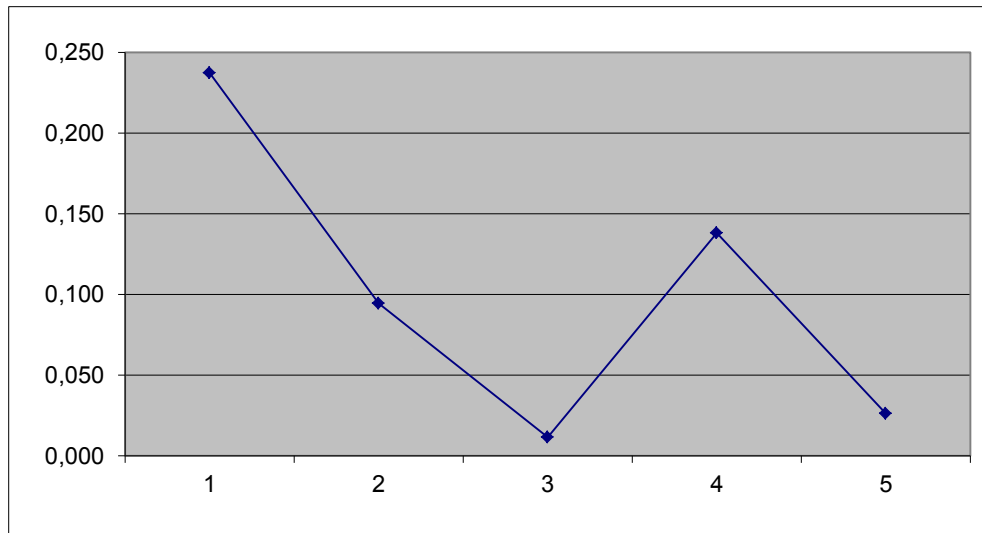
Мал. 3.2.5. Коефіцієнт загальної ліквідності



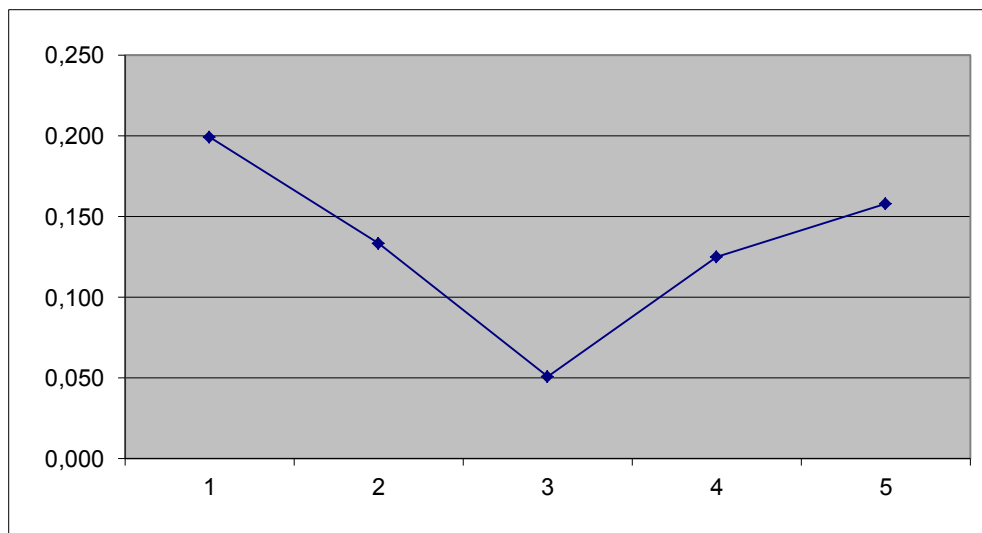
Мал. 3.2.6. Коефіцієнт поточної ліквідності



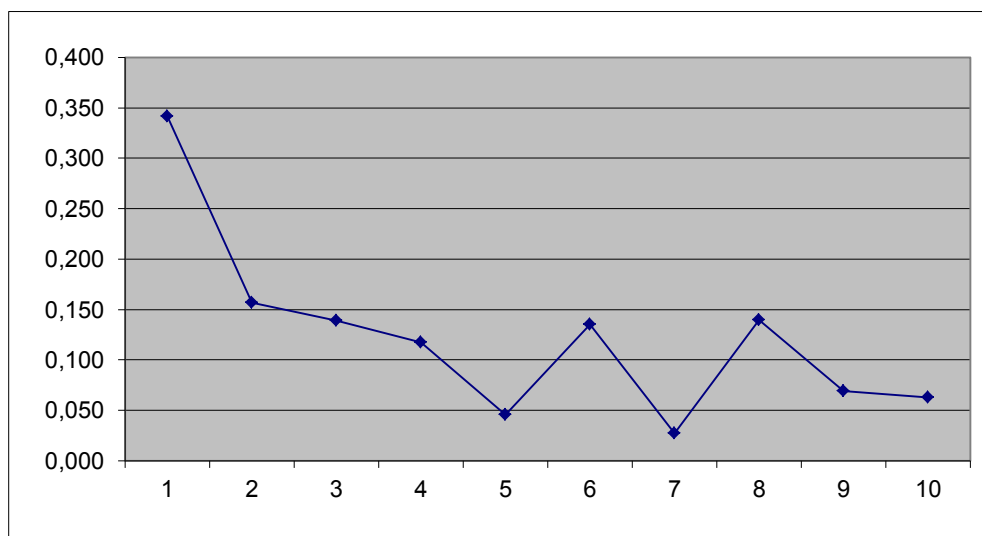
Мал. 3.2.7. Коефіцієнт миттєвої ліквідності



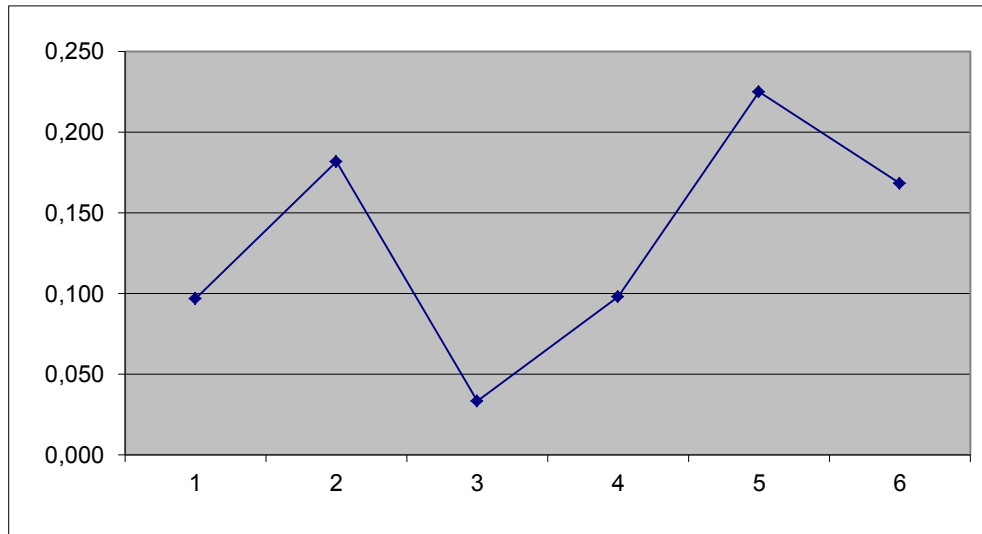
Мал. 3.2.8. Коефіцієнт незалежності



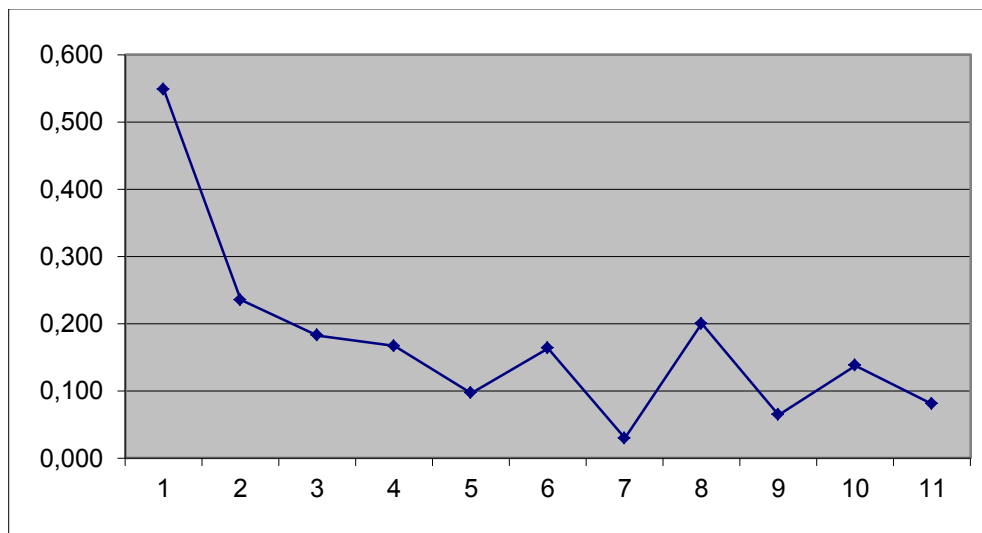
Мал. 3.2.9. Коефіцієнт ділової активності



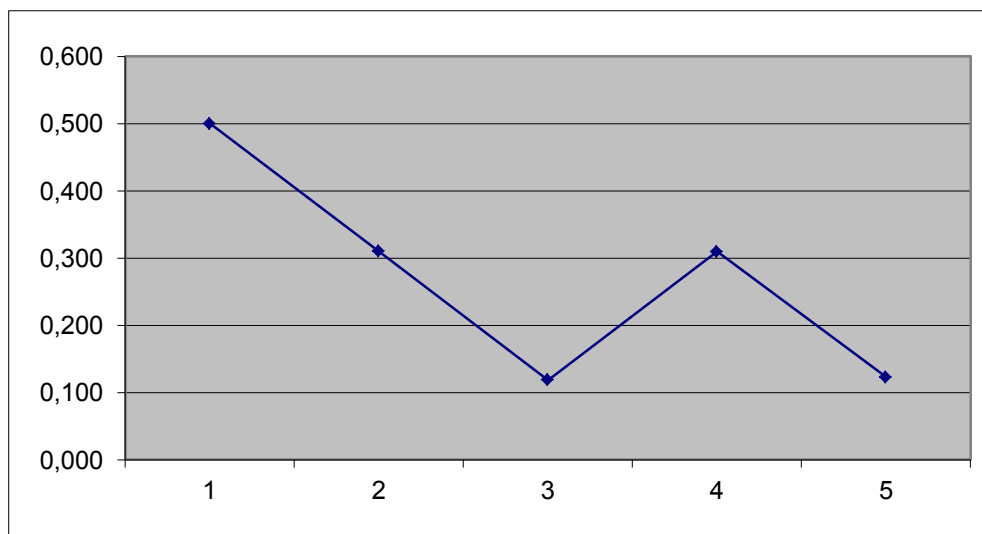
Мал. 3.2.10. Коефіцієнт маневреності власних коштів



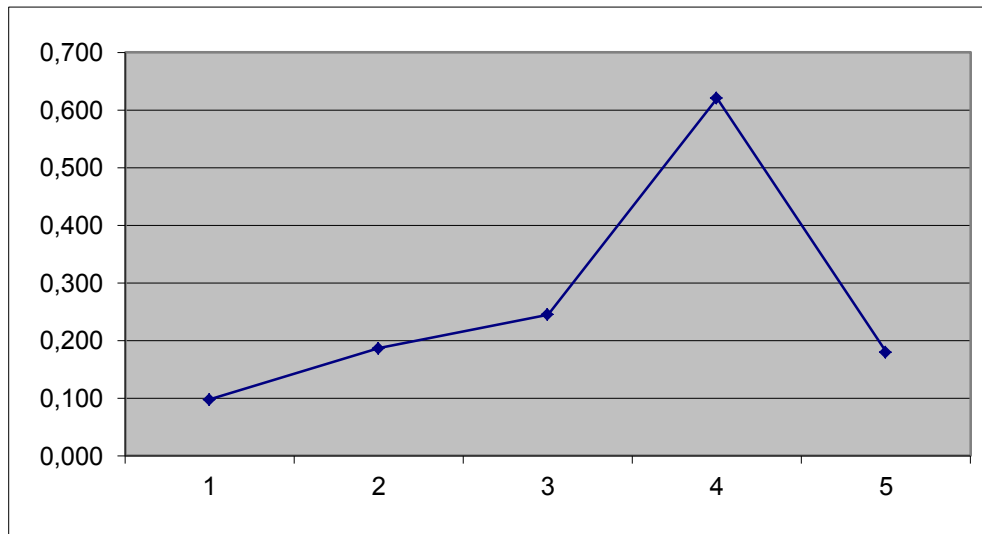
Мал. 3.2.11. Коефіцієнт рентабельності активів



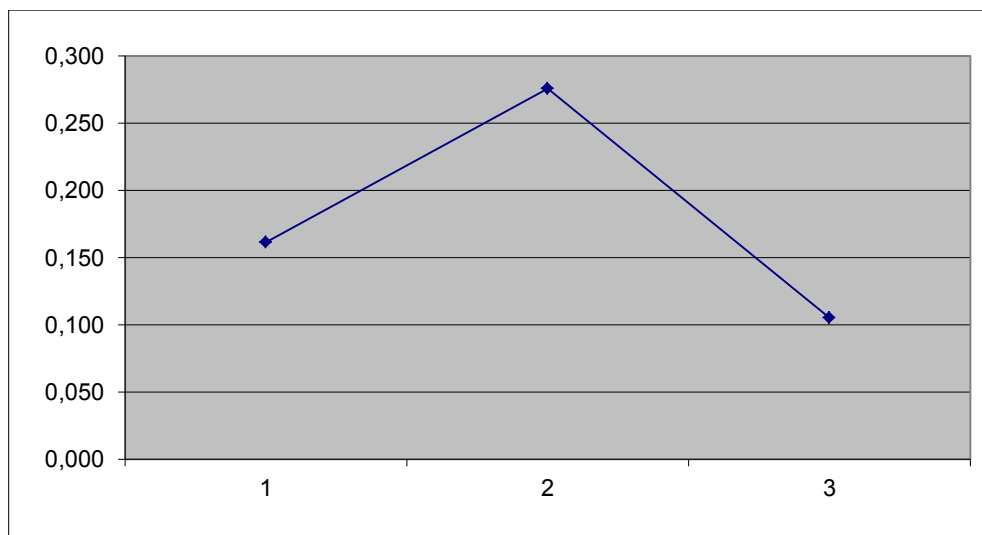
Мал. 3.2.12. Коефіцієнт рентабельності продаж



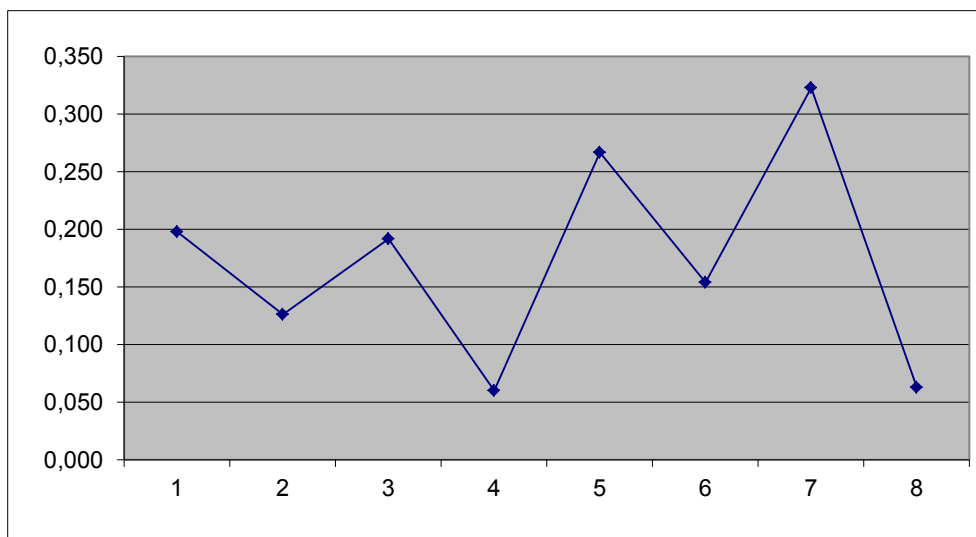
Мал. 3.2.13. Період обіговості оборотних активів



Мал. 3.2.14. Коефіцієнт достатності робочого капіталу

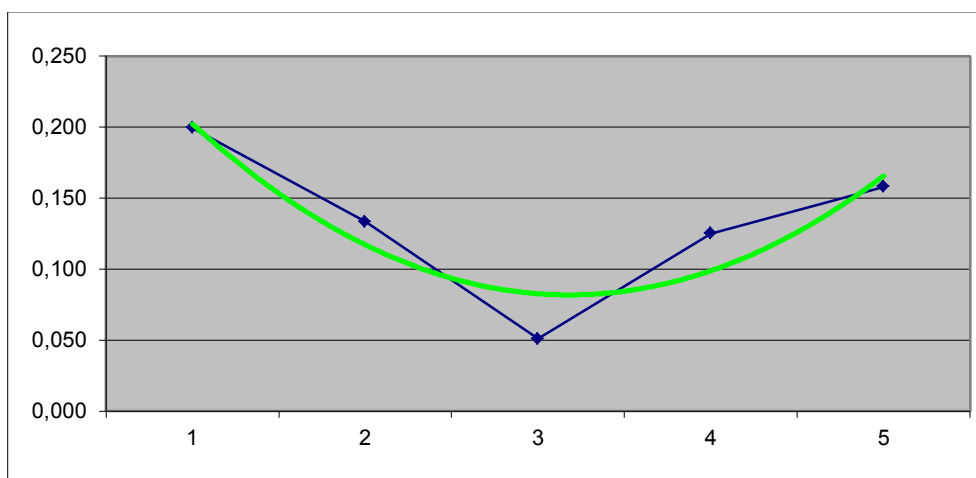


Мал. 3.2.15. Коефіцієнт покриття зобов'язань EBITDA



Як бачимо з графіків, кредитоспроможність позичальника дійсно в більшості випадків не залежить монотонно від значень коефіцієнтів. В деяких випадках часті „стрибки” важко пояснити з економічної точки зору. Проте варто нагадати, що використовуються статистичні дані і для покращення самої моделі потрібно зробити щонайменше з 10 експериментів на основі різних вибірок. Також можна апроксимувати отримані результати (наприклад, використовуючи логістичну або поліноміальну апроксимацію), встановлюючи таким чином чіткішу залежність між значеннями коефіцієнтів і кредитоспроможністю позичальника. Наприклад, для коефіцієнта незалежності матимемо:

Мал. 3.2.16. Коефіцієнт незалежності (апроксимація)



3.3. Побудова моделі

Використовуючи отримані бальні оцінки фінансових коефіцієнтів позичальника, побудуємо модель. Для побудови логістичної регресії використаємо пакет EViews3.1. Матимемо:

Dependent Variable: DEF

Method: ML - Binary Logit

Date: 10/28/08 Time: 10:29

Sample: 1 562

Included observations: 562

Convergence achieved after 6 iterations

Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
K1 - коефіцієнт загальної ліквідності	4.812120	1.293175	3.721166	0.0002
K2 – коеф. поточної ліквідності	3.936897	0.940689	4.185119	0.0000
K3 – коеф. миттєвої ліквідності	6.722845	2.176591	3.088703	0.0020
K4 – коеф. фінанс. незалежності	6.244207	3.771729	1.655529	0.0978
K5 – коеф. ділової активності	0.307719	1.691504	0.181921	0.8556
K6 – коеф. маневреності вл. коштів	8.832650	3.648220	2.421084	0.0155
K7 – рентабельність активів	3.759479	1.670388	2.250662	0.0244
K8 – рентабельність продажу	1.084062	1.724517	0.628618	0.5296
K9 - період обіговості оборот. активів	4.502230	1.097351	4.102816	0.0000
K10 - коеф. достатності роб. капіталу	12.02321	4.526691	2.656070	0.0079
K11 – коеф. покриття зоб. EBITDA	5.013296	2.280268	2.198555	0.0279
C	-11.79819	1.452816	-8.120911	0.0000
Mean dependent var	0.163701	S.D. dependent var		0.370334
S.E. of regression	0.294306	Akaike info criterion		0.602215
Sum squared resid	47.63871	Schwarz criterion		0.694703
Log likelihood	-157.2225	Hannan-Quinn crit.		0.638324
Restr. log likelihood	-250.5151	Avg. log likelihood		-0.279755
LR statistic (11 df)	186.5852	McFadden R-squared		0.372403
Probability(LR stat)	0.000000			
Obs with Dep=0	470	Total obs		562
Obs with Dep=1	92			

AUC = 0,8904

Максимізуючи критерій (2.2.4.6), отримаємо:

Cut_off = 0,1098 – значення порогу відсікання;

Se+Sp = 1,6340 – критерій оптимізації.

Для оцінки впливу попередньої обробки даних на якість моделі побудуємо модель на основі реальних значень фінансових коефіцієнтів. Отримаємо:

Dependent Variable: DEF

Method: ML - Binary Logit

Date: 10/17/08 Time: 10:16

Sample: 1 562

Included observations: 562

Convergence achieved after 6 iterations

Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
K1 - коефіцієнт загальної ліквідності	-0.402756	0.318108	-1.266099	0.2055
K2 – коеф. поточної ліквідності	-0.850248	0.344179	-2.470369	0.0135
K3 – коеф. миттєвої ліквідності	-2.011100	0.854425	-2.353747	0.0186
K4 – коеф. фінанс. незалежності	-0.291257	0.303355	-0.960121	0.3370
K5 – коеф. ділової активності	-0.352293	0.216505	-1.627181	0.1037
K6 – коеф. маневреності вл. коштів	0.903037	0.431270	2.093900	0.0363
K7 – рентабельність активів	-0.679391	0.204759	-3.318009	0.0009
K8 – рентабельність продажу	0.221232	0.169088	1.308384	0.1907
K9 - період обіговості оборот. активів	0.128754	0.175619	0.733145	0.4635
K10 - коеф. достатності роб. капіталу	-0.053842	0.199060	-0.270483	0.7868
K11 – коеф. покриття зоб. EBITDA	0.672825	0.232667	2.891795	0.0038
C	-2.439576	0.236531	-10.31398	0.0000
Mean dependent var	0.163701	S.D. dependent var		0.370334
S.E. of regression	0.338660	Akaike info criterion		0.793551
Sum squared resid	63.07980	Schwarz criterion		0.886039
Log likelihood	-210.9879	Hannan-Quinn criter.		0.829659
Restr. log likelihood	-250.5151	Avg. log likelihood		-0.375423
LR statistic (11 df)	79.05453	McFadden R-squared		0.157784
Probability(LR stat)	2.25E-12			
Obs with Dep=0	470	Total obs		562
Obs with Dep=1	92			92

AUC = 0,7858

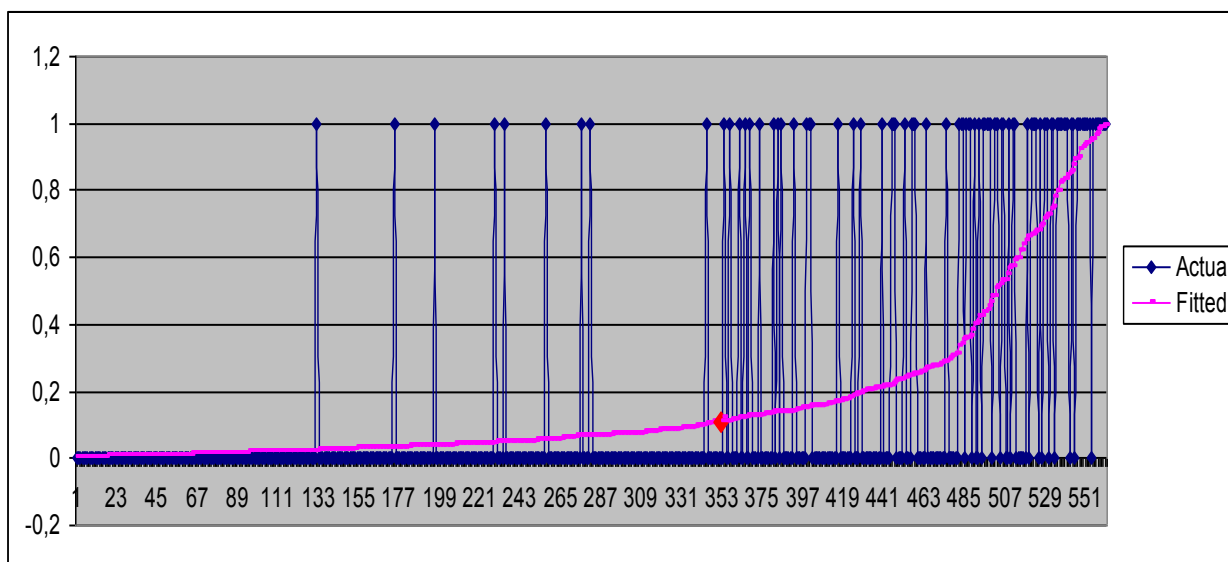
Cut_off = 0,2275 – значення порогу відсікання;

Se+Sp = 1,4713 – критерій оптимізації.

Як бачимо попередня обробка значно покращує якість моделі та її передбачувальну здатність. Значення AUC першої моделі більше за AUC другої, що свідчить про кращу прогностичну силу моделі, побудованої відповідно до бальних оцінок коефіцієнтів на отриманих інтервалах. Приймаючи до уваги значення критеріїв Акайке і Шварца, можна констатувати вищу адекватність першої моделі (значення критеріїв першої моделі менше за відповідні значення другої). Також згідно отриманих значень критеріїв оптимізації (максимальної сумарної чутливості та

специфічності) можна говорити про вищу розподільну здатність першої моделі, оскільки значення критерію цієї моделі приймає більше значення. Продemonструємо це за допомогою графіків реальних (Actual) та отриманих (Fitted) значень побудованих моделей. Для першої моделі маємо:

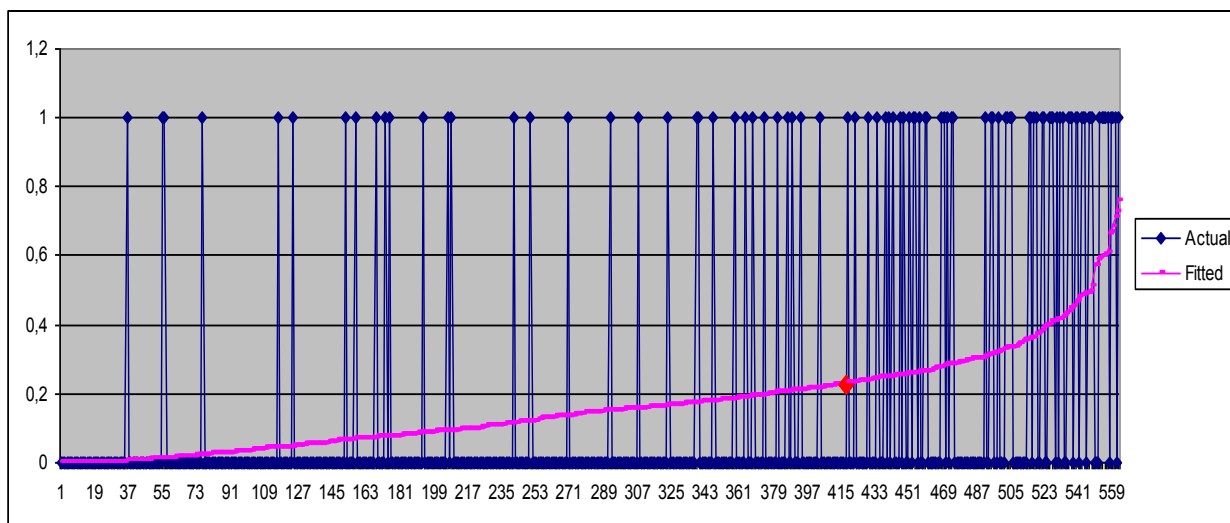
Мал. 3.3.1. Графік реальних та отриманих значень 1-ої моделі



Точка на графіку отриманих значень вказує на поріг відсікання, тобто при отриманні результату ймовірності дефолту вище даного порогу клієнту відмовляють у видачі кредиту (його фінансове становище вважається незадовільним, а сам клієнт – не кредитоспроможним). З графіку видно, що модель може доволі чітко відрізнити кредитоспроможного клієнта від такого, що зазнав дефолту.

Побудуємо аналогічний графік для другої моделі:

Мал. 3.3.2. Графік реальних та отриманих значень 2-ої моделі



Аналізуючи його, можна сказати про низьку розподільну здатність другої моделі. Частина невірно спрогнозованих значень, менших порогу відсікання, (помилки першого роду) є досить великою.

Отже, провівши аналіз побудованих моделей можна чітко констатувати, що модель, побудована на основі попередньо оброблених даних з використанням запропонованого методу, за всіма показниками (прогностична сила, адекватність, розподільна здатність) є кращою ніж модель, побудована на реальних значеннях вхідних параметрів. Можемо зробити висновок, що запропонований метод попередньої обробки даних значно підвищує якість моделі та допомагає отримати непогані результати.

РОЗДІЛ 4

РОЗРОБКА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

4.1. Функціональність програмного забезпечення

Для реалізації присвоєння бальних оцінок фінансовим коефіцієнтам позичальника банку було розроблене програмне забезпечення. Зокрема був реалізований алгоритм розбиття на інтервали з визначенням відповідних значень ймовірностей дефолту на них, наведений у розділі 2.3.2.

В програмі є можливість обчислення інтервалів для кожного окремого коефіцієнта з виведенням проміжних результатів, а також обчислення та виведення інтервалів для всіх запропонованих користувачем коефіцієнтів відразу. Швидкість роботи програми збільшена завдяки використанню рекурсії (наступні значення ймовірностей дефолту обчислюється на основі попередніх).

Програма була реалізована на мові програмування Visual Basic. Програма має наступні обмеження:

- максимальна кількість позичальників – 10000;
- максимальна кількість параметрів позичальників – 50;
- максимальна кількість значень кожного параметру – 10000.

Вхідними даними в програмі є:

- код ЄДРПОУ позичальника;
- дефолт/недефолт позичальника на останню дату;
- фінансові коефіцієнти позичальника.

Вихідними даними є:

- інтервали з відповідними ймовірностями дефолту (бальні оцінки), розраховані для кожного окремого фінансового коефіцієнта позичальника.
- перелік всіх позичальників з розрахованими бальними оцінками для його коефіцієнтів.

4.2. Опис інтерфейсу

Вікно програми складається з семи закладок:

- Коеф_ост_дата – закладка для введення сукупності фінансових коефіцієнтів позичальників на останню звітну дату, дані можуть завантажуватись з бази даних Microsoft Access;

Мал. 4.2.1. Закладка „Коеф_ост_дата”

Дефолт	Код ЄДРПОУ	Дата	Коефіцієнт загальної ліквідності	Коефіцієнт поточної ліквідності	Коефіцієнт миттєвої ліквідності	Коефіцієнт фінансової стійкості	Коефіцієнт незалежності
0	342	2008-07-01	1.02940	0.64200	-0.13103	0.07048	13.18892
0	343	2008-07-01	1.19310	1.19310	1.08847	0.27097	2.69047
0	344	2008-07-01	1.12783	0.82395	0.00000	0.11347	7.81314
0	345	2008-07-01	2.35862	0.94287	0.06403	0.66179	43.80089
0	346	2008-07-01	1.74657	0.86273	0.04926	0.70007	1.52666
0	347	2008-07-01	0.91228	0.55167	0.01263	0.10536	8.49114
0	348	2008-07-01	0.41878	0.33867	0.00617	0.57469	3.08945
0	349	2008-07-01	1.00943	0.48414	0.05187	0.02152	45.47236
0	350	2008-07-01	0.84455	0.29363	0.03368	0.06729	13.86032
0	351	2008-07-01	0.92952	0.34986	0.03162	0.31175	6.01065
0	352	2008-07-01	1.15610	1.15559	0.44762	0.25204	2.96767
0	353	2008-07-01	0.39227	0.31897	0.00000	0.70358	0.75796
0	354	2008-07-01	1.24913	0.92632	0.16274	0.92260	0.51587
0	355	2008-07-01	1.22950	0.58149	0.02613	0.39879	39079.19355
0	356	2008-07-01	1.30306	0.86817	0.12559	0.45897	9.18997
0	357	2008-07-01	1.67944	1.07162	0.38911	0.48023	1.09597
0	358	2008-07-01	7.80719	7.80424	7.15431	0.87494	0.28747
0	359	2008-07-01	6.81540	6.66062	0.00224	0.97586	154.40913
0	360	2008-07-01	2.38458	1.45587	0.00787	0.63635	2.99338
0	361	2008-07-01	3.03365	0.88445	0.42972	0.68777	0.45397
0	362	2008-07-01	1.45224	1.12291	0.00187	0.33703	3.92470

- Коеф_без_пустих – закладка, де користувач має змогу позбутися некоректних та граничних значень коефіцієнтів, що негативно впливають на розподіл, а також провести розрахунок математичного сподівання та дисперсії для відібраних значень.

Мал. 4.2.2. Закладка „Коеф_без_пустих”

Дефолт	Код ЄДРПОУ	Коефіцієнт загальної ліквідності K1	Коефіцієнт поточної ліквідності K2	Коефіцієнт миттєвої ліквідності K3	Коефіцієнт незалежності K4	Коефіцієнт ділової активності K5	Коефіцієнт маневреності власних коштів K6
0	541	1.00140	0.85304	0.00000	385.26488	0.41705	-0.11734
0	542	2.03011	1.77825	0.05142	6.52702	2.00863	0.56415
0	543	1.47927	1.14990	0.02775	21.16628	4.14618	0.62413
0	544	1.27672	0.62177	0.00330	11.04627	2.71263	0.11538
0	545	0.36068	0.35699	0.08115	12.98719	0.31781	-8.89359
0	546	0.68811	0.46533	0.06197	4.17482	3.51771	-1.37745
0	547	0.99777	0.71290	0.01128	68.50656	5.99602	0.27223
0	548	1.04416	0.60018	0.00000	6.81116	10.69500	0.30128
0	549	3.45032	3.11661	0.36958	0.47399	3.27861	0.99508
0	550	1.76250	1.58005	0.15187	1.36152	11.64578	0.97085
0	551	1.96986	0.28970	0.00060	35.97639	3.62594	-0.55292
0	552	0.93072	0.73868	0.03686	1.84252	1.95317	-1.11040
0	553	5.56976	1.14843	0.01163	0.14322	15.18198	0.99668
0	554	0.25796	0.25786	0.00377	102.67458	0.38351	-76.18892
0	555	7.32029	3.38450	0.53707	0.15263	2.10568	0.39906
0	556	2.42857	1.59987	0.00689	1.02784	2.39137	0.56602
0	557	1.07286	0.76269	0.01931	7.79042	5.77406	0.75947
0	558	0.97182	0.40522	0.05020	13.03744	22.61962	-0.19231
0	559	0.29290	0.06193	0.00315	1.67944	0.67515	-1.18763
0	560	1.00016	0.07702	0.02611	362.25959	0.29019	0.21386
0	561	0.78687	0.77384	0.00232	17.27562	0.94807	-11.70512
0	562	1.08743	0.68930	0.00000	10.43195	10.17169	0.97028
	середнє	1.66768	0.95614	0.11364	12.61688	3.65857	-1.55544
	дисперсія	1.65595	0.87869	0.14835	1602.28850	27.19568	47.50992

- Коеф_ранж – закладка, на якій проводиться стандартизація даних шляхом віднімання середнього та ділення на квадрат дисперсії. Таким чином користувач має нагоду врівноважити шкалу значень для різних коефіцієнтів.

Мал. 4.2.3. Закладка „Коеф_ранж”

Дефолт	Код ЄДРПОУ	Коефіцієнт загальної ліквідності K1	Коефіцієнт поточної ліквідності K2	Коефіцієнт миттєвої ліквідності K3	Коефіцієнт незалежності K4	Коефіцієнт ділової активності K5	Коефіцієнт маневреності власних коштів K6
1	1	-0.47817	-0.06056	-0.21534	-0.30653	-0.61361	0.21896
1	2	-0.30548	-0.21608	-0.27891	-0.29421	-0.66826	0.16695
1	3	-1.15100	-0.92788	-0.27711	-0.30206	-0.66370	0.16347
1	4	-0.33908	-0.53875	-0.28360	-0.30737	-0.66521	0.21806
1	5	-0.56446	-0.67332	-0.29178	-0.30656	-0.41245	0.22272
1	6	-0.84149	-0.57776	-0.29424	-0.22456	-0.30883	-0.03383
1	7	-0.86206	-0.96723	-0.24121	-0.01559	-0.16901	-0.55471
1	8	-0.50195	-0.33651	-0.28862	-0.14561	-0.59180	0.24609
1	9	-0.46092	-0.77857	-0.09690	-0.30875	-0.64303	0.19704
1	10	-1.09169	-0.87913	-0.27211	-0.29608	-0.64163	0.15232
1	11	0.74826	1.69581	-0.27114	-0.31316	-0.65509	0.24493
1	12	-1.20475	-0.99163	-0.29432	-0.06506	-0.69787	-1.05759
1	13	-1.16123	-0.96221	-0.29504	-0.30285	-0.69694	0.16637
1	14	-0.64773	-0.84871	-0.29063	-0.30591	-0.68563	0.21555
1	15	-0.33990	-0.94640	-0.25386	-0.31044	-0.65851	0.22394
1	16	-0.82111	-0.81514	-0.29486	-0.29718	-0.66205	0.17854
1	17	-0.47311	-0.70129	-0.26801	-0.30615	-0.66474	0.22517
1	18	0.62249	0.04729	-0.29414	-0.30824	-0.63384	0.28524
1	19	0.26225	-0.50676	-0.28234	-0.30524	-0.60541	0.28412
1	20	1.32033	2.57163	0.36734	-0.31301	-0.66458	0.31456
1	21	0.96904	-0.12303	0.25002	-0.31078	-0.64679	0.27480
1	22	-1.11926	-0.92783	-0.28312	-0.27452	-0.66851	0.02833

- Аналіз_інтервалів – закладка, на якій користувач має змогу провести розрахунок бальних оцінок з розбиттям на інтервали для обраного фінансового коефіцієнта з виведенням проміжних результатів. Виводиться відсортована вибірка значень коефіцієнтів із значеннями дефолту/недефолту позичальника та відмічаються кольором границі отриманих інтервалів. Для запуску розрахунку необхідно натиснути кнопку „Розрахувати інтервали” на цій закладці.

Мал. 4.2.4. Закладка „Аналіз_інтервалів”

розрахованих інтервалів. Для виведення бальних оцінок необхідно натиснути „Вивести” на цій закладці.

Мал. 4.2.6. Закладка „Коеф_ймовірн”

Вивести		Коефіцієнт загальної ліквідності K1	Коефіцієнт поточної ліквідності K2	Коефіцієнт миттєвої ліквідності K3	Коефіцієнт незалежності K4	Коефіцієнт ділової активності K5	Коефіцієнт маневреності власних коштів K6
Дефолт	Код ЄДРПОУ						
0	175260	0.22523	0.22727	0.23724	0.19937	0.34146	0.22500
0	307371	0.15385	0.18605	0.23724	0.19937	0.34146	0.22500
0	308844	0.48837	0.64706	0.23724	0.19937	0.34146	0.22500
0	413133	0.15385	0.24444	0.23724	0.19937	0.34146	0.22500
0	991841	0.12698	0.08333	0.23724	0.19937	0.15646	0.22500
0	1033303	0.09091	0.24444	0.23724	0.19937	0.15646	0.09804
0	1527940	0.48837	0.64706	0.23724	0.05085	0.11765	0.18182
0	1973517	0.22523	0.05556	0.23724	0.13333	0.15646	0.16835
0	3302003	0.22523	0.35484	0.09459	0.19937	0.34146	0.22500
0	3328675	0.48837	0.64706	0.23724	0.19937	0.34146	0.22500
0	3350137	0.12500	0.05882	0.23724	0.19937	0.34146	0.16835
0	3748437	0.48837	0.64706	0.23724	0.05085	0.34146	0.09677
0	3750635	0.48837	0.64706	0.23724	0.19937	0.34146	0.22500
0	3765967	0.12698	0.64706	0.23724	0.19937	0.34146	0.22500
0	3802027	0.15385	0.64706	0.23724	0.19937	0.34146	0.22500
0	4594686	0.09091	0.35484	0.23724	0.19937	0.34146	0.22500
0	5278266	0.22523	0.08333	0.23724	0.19937	0.34146	0.22500
0	5446858	0.12500	0.22727	0.23724	0.19937	0.34146	0.16835
0	5516470	0.25000	0.24444	0.23724	0.19937	0.34146	0.16835
0	9322018	0.11429	0.05882	0.13793	0.19937	0.34146	0.16835
0	9794409	0.11429	0.07353	0.13793	0.19937	0.34146	0.16835
0	13525460	0.48837	0.64706	0.23724	0.19937	0.34146	0.09804

- Аналіз_інтервалів_реальні_знач – закладка, на якій користувач має змогу повернутися до реальних значень границь інтервалів шляхом операції зворотній стандартизації (множення на корінь дисперсії та додавання середнього).

Мал. 4.2.6. Закладка „Коеф_ймовірн”

Коефіцієнт загальної ліквідності K1			Коефіцієнт поточної ліквідності K2			Коефіцієнт миттєвої ліквідності K3			Коефіцієнт незалежності K4		
ліва грань	права грань	Ймовір ність	ліва грань	права грань	Ймовір ність	ліва грань	права грань	Ймовір ність	ліва грань	права грань	Ймові рність
0.046677	0.565966	0.488	0.004698	0.16338	0.647	0	0.038686	0.237	0.059569	4.26257	0.199
0.565966	0.811477	0.091	0.16338	0.257861	0.355	0.038686	0.080357	0.095	4.26257	8.291225	0.133
0.811477	0.991223	0.127	0.257861	0.369897	0.083	0.080357	0.208815	0.011	8.291225	13.86032	0.051
0.991223	1.140898	0.225	0.369897	0.481109	0.244	0.208815	0.346299	0.138	13.86032	24.99326	0.125
1.140898	1.27458	0.154	0.481109	0.575065	0.143	0.346299	7.154313	0.026	24.99326	548.8199	0.158
1.27458	1.448688	0.083	0.575065	0.672555	0.056						
1.448688	1.577442	0.207	0.672555	0.7722	0.186						
1.577442	1.876811	0.022	0.7722	0.886609	0.074						
1.876811	2.126282	0.250	0.886609	1.000465	0.227						
2.126282	2.630571	0.125	1.000465	1.149902	0.026						
2.630571	3.575576	0.114	1.149902	1.408503	0.207						
3.575576	9.957313	0.000	1.408503	1.955875	0.026						
			1.955875	7.804237	0.059						

Програма має зручний інтерфейс і є простою у використанні. Всі розрахунки проводяться достатньо швидко і не потребують значних ресурсів ПК.

РОЗДІЛ 5

РОЗРОБЛЕННЯ СТАРТАП ПРОЕКТУ

5.1 Опис ідеї

Таблиця 5.1 Опис ідеї стартап-проекту

Зміст ідеї	Напрямки застосування	Вигоди для користувача
Оцінка платоспроможності клієнта на певну дату, і прогноз фінансової стабільності на перспективу, на основі регресійної моделі.	1. Використання в банках.	1. Забезпечення фінансової стабільності. 2. Мінімізація ризиків капіталовкладень.
	2. Використання в мережах ломбардів, як окремого виду бізнеса.	

Таблиця 5.2 Визначення сильних, слабких та нейтральних характеристик ідеї проекту

п/п	Техніко-економічні характеристики ідеї	концепції конкурентів				W (слабка сторона)	N (нейтральна сторона)	S (сильна сторона)
		Crewor	https://www.standardandpoors.com/en_US/web/guest/home	https://www.moodys.com/researchandratings/	https://www.creditkarma.com			
1	Призначення, Економічні - Business value, вартість продукту	Вартість продукту і в його унікальності, автоматизації	Підкріплення існуючих марки	Престиж створеної скорингової моделі	Багатофункціональний сервіс та калькулятор	Гнучкість слабка на початкових етапах	Для малих фірм незначна, проте як і для конкурентів	Значний об'єм функціоналу
2	Економічні - Ширина покриття ринку	Банки та крупні позичальники,	Банки та фірми-позичальники	Банки та фірми-позичальники,	Більші за рахунок коштування	Важко відбити банки у їхнього ж	Доступ до портфелю може бути	В майбутньому можна розробити

		що мають доступ до кредитного портфелю		клієнти банків		керівництва	відсутні. Потребуватиметься створення локального, проте це означає наповнення історії кредитів.	ти адміністративний модуль в якому будуть варіанти кастомізації
3	Технологічні Економічні - Наявність людського втручання в модель	Відсутня наявність людського втручання в модель	Аналіз проводиться людьми	Аналіз проводиться людьми	Зведена до мінімуму, проте наявна	Чим менше вартість продукту тим менше факторів враховується	Для маленьких та близьких до середніх банків переваги й мінуси незначні	Коефіцієнт ціна/прибуток більший

5.2 Технологічний аудит ідеї проекту

Таблиця 5.3 Технологічна здійсненність ідеї проекту

п/п	Ідея проекту	Технології її реалізації	Наявність технологій	Доступність технологій
	Створення ПЗ на основі логічної регресійної моделі для оцінки кредитоспроможності клієнта банку	Використання лінійних моделей або таких, що приводяться до лінійного виду шляхом перетворення змінних, наприклад логарифмування.	Наявна технологія	Доступна
		Логічна регресійна модель	Наявна технологія	Доступна
<p>Обрана технологія реалізації ідеї проекту: Лінійна модель безперечно, містить у собі певну умовність, оскільки передбачає однаковий характер зв'язку з усіма факторами. Проте використання надто складних функцій неминує веде до збільшення кількості параметрів, а отже, зменшує точність вимірювання та ускладнює інтерпретацію результатів. Тому доцільніше обрати регресійну модель. З математичної точки зору, це достатньо гнучка функція, та, по-друге, вона дає легку для сприйняття інтерпретацію.</p>				

5.3 Аналіз ринкових можливостей запуску стартап-проекту

Визначимо ринкові можливості, які можна використати під час ринкового впровадження проекту, та ринкових загроз, які можуть перешкодити реалізації проекту

Таблиця 5.4 Попередня характеристика потенційного ринку стартап-проекту

п/п	Показники стану ринку (найменування)	Характеристика
	Кількість головних гравців, од	7
	Загальний обсяг продаж, грн/ум.од	2500000
	Динаміка ринку (якісна оцінка)	Зростає/спадає/стагну є
	Наявність обмежень для входу (вказати характер обмежень)	Здобуття перших банків
	Специфічні вимоги до стандартизації та сертифікації	Оформлення документації
	Середня норма рентабельності в галузі (або по ринку), %	12%

Також визначимо потенційні групи клієнтів, їх характеристики, та сформуємо орієнтовний перелік вимог до товару для кожної групи

Таблиця 5.5 Характеристика потенційних клієнтів стартап-проекту

п/п	Потреба, що формує ринок	Цільова аудиторія (цільові сегменти ринку)	Відмінності у поведінці різних потенційних цільових груп клієнтів	Вимоги споживачів до товару
	Об'єктивна оцінка фінансової стабільності позичальника й облік можливих ризиків по кредитних операціях дозволяють банку ефективно управляти кредитними ресурсами й отримувати прибуток	Основними потенційними клієнтами є банки і ломбарди.	Банки можуть проводити порівняння статистики повернених коштів і не повернених за деякі періоди своєї діяльності (до використання нашого продукту та після). Ломбарди ж зазвичай не мають такої великої історії кредитів, тому вони можуть дивитись на	- точність прогнозів; - якість служби підтримки; - зручний інтерфейс; - продуманий UX.

			рівень свого доходу	
--	--	--	---------------------	--

Після визначення потенційних груп клієнтів проведемо аналіз ринкового середовища: складемо таблиці факторів, що сприяють ринковому впровадженню проекту, та факторів, що йому перешкоджають (табл. № 6-7)

Таблиця 5.6 Фактори загроз

п/п	Фактор	Зміст загрози	Можлива реакція компанії
	Робота з банками, що веде за собою бюрократію	Керівництво банків та ломбардів не завжди готові до нового та неперевіреного. Вони не люблять ризикувати із нововведеннями. Щоб бути реалізованою, новій ідеї необхідно пройти дуже багато бюрократичних етапів. Банки відмовлять нам у співробітництві через недостатність досвіду та новий підхід.	Компанія може надати докази підвищення прибутку та ефективності в цифрах.
	Ломбарди можуть бути незацікавленими.	Ломбарди зазвичай мають менший грошовий обіг, ніж банки. Вони відмовлять у співробітництві, тому що будуть не зацікавлені у всіх можливостях нашого продукту.	Компанія може придумати бізнес-модель, що включає в себе пробний термін.

Таблиця 5.7 Фактори можливостей

п/п	Фактор	Зміст можливості	Можлива реакція компанії
	Тотальна диджиталізація	скорочення кількості паперових бумаг та клерків, що їх перебирають.	Ввести рекламну компанію про користь процесу автоматизації аналізу кредитоспроможності
	Фінансова криза	Банк зрозуміє нагальну потребу більш раціонально	В ІТ сфері гроші будуть при фінансовій кризі, тому гроші від інвесторів будуть

		розділяти свої поточні кошти	
--	--	------------------------------	--

Визначимо загальні риси конкуренції на ринку (табл. 8).

Таблиця 5.8 Ступеневий аналіз конкуренції на ринку

Особливості конкурентного середовища	В чому проявляється дана характеристика	Вплив на діяльність підприємства (можливі дії компанії, щоб бути конкурентоспроможною)
1. Тип конкуренції - олігополія	На ринку є декілька інших компаній, що пропонують схожий продукт, як і наш.	Зниження цін, використання реклами, поліпшення якості
2. За рівнем конкурентної боротьби - національний	Наш продукт можна використовувати для будь-якого банку України, оскільки політика банків залишається однією ж.	Тримати руку на пульсі нововведень зі сторони законів країни.
3. За галузевою ознакою - внутрішньогалузева	Наш продукт застосовується лише в одній галузі.	Впровадження нових технологій для зменшення витрат.
4. Конкуренція за видами товарів: - товарно-видова	Різновиди однієї категорії товару, які здатні задовольнити конкретне бажання покупця.	Робота над індивідуальними підходами до кожного клієнта
5. За характером конкурентних переваг - нецінова	Основною перевагою є досягнення результату більш простим шляхом	Вдосконалення технологій.
6. За інтенсивністю - не марочна	Поки що немає необхідності в присвоєнні марки.	На даному етапі не так важливо дотримуватись марочної політики.

Проведемо аналіз конкуренції за М. Портером.

Таблиця 5.9 Аналіз конкуренції в галузі за М. Портером

	Прямі конкуренти в галузі	Потенційні конкуренти	Постачальники	Клієнти	Товари-замінники
Складові аналізу	Компанія Moody's та S&P Global	Барерами є: - Розмір капіталовкладень. Доступ до ресурсів.	Основним постачальником є інвестор. І основним фактором сили є кількість інвестицій.	Факторами сили споживача є вимоги до продукту, та чутливості до змін цін	немає

Висновки :	За інтенсивністю – марочна.	- можливість виходу на ринок є - потенційні конкуренти є.	Чим більше інвестор готовий витратити, тим краще може бути продукт, і навпаки.	Клієнти диктують умови роботи на ринку. Від їх потреб та забаганок може виникнути необхідність у вдосконаленні продукту	немає
------------	-----------------------------	--	--	---	-------

На основі аналізу конкуренції, проведеного в п. 3.5 (табл. 9), а також із урахуванням характеристик ідеї проекту (табл. 2), вимог споживачів до товару (табл. 5) та факторів маркетингового середовища (табл. №№ 6-7) визначемо перелік факторів конкурентоспроможності.

Таблиця 5.10 Обґрунтування факторів конкурентоспроможності

п/п	Фактор конкурентоспроможності	Обґрунтування (наведення чинників, що роблять фактор для порівняння конкурентних проектів значущим)
	Рівень технологій	Інші фірми не використовують автоматичне визначення кредитоспроможності. Наш продукт на таке здатен.
	Надійність	Ми будемо підтримувати наш продукт на той етап, на який погодиться клієнт. Це буде гарантувати надійність та «ремонтпридатність» продукту.
	Естетичність	Наш продукт буде мати гарний інтерфейс, в якому буде зручно працювати.

Проводимо аналіз сильних та слабких сторін стартап-проекту (табл. 11).

Таблиця 5.11 Порівняльний аналіз сильних та слабких сторін «Crewor»

п/п	Фактор конкурентоспроможності	Бали 1-20	Рейтинг товарів-конкурентів у порівнянні з Crewor						
			3	2	1		1	2	3
	Рівень технологій	8							

	Надійність	7							
	Естетичність	10							

Таблиця 5.12 SWOT- аналіз стартап-проекту

<p>Сильні сторони:</p> <ul style="list-style-type: none"> - Вартість підтримки та послідовних замовлень менше ніж у конкурентів - За наявністю великої кількості даних точність зростає - Залучення інвесторів та розвиток функціоналу 	<p>Слабкі сторони:</p> <ul style="list-style-type: none"> - Вартість першої одиниці продукту дуже висока - Банки, ломбарди та інші позичальники можуть бути незацікавленими в не провірній програмі. - Відсутність початкових клієнтів та їх відгуків - Невідомість моделі у сфері банків на рівні конкурентів.
<p>Можливості:</p> <ul style="list-style-type: none"> - Гра на тотальній глобалізації та оптимізації сфери банківської справи. - Створити продукт з гнучким функціоналом та можливістю його доповнення - Відсутність помилок, створених людським фактором. 	<p>Загрози:</p> <ul style="list-style-type: none"> - Робота з банками, що веде за собою бюрократію - Ломбарди можуть бути незацікавленими. - Бар'єр для виходу на ринок збільшується як плине час

Таблиця 5.13 Альтернативи ринкового впровадження стартап-проекту

п/п	Альтернатива (орієнтовний комплекс заходів) ринкової поведінки	Ймовірність отримання ресурсів	Строки реалізації
	Перетворення бізнес-моделі у формат щомісячної підписки	95%	Перетворення внутрішньої структури платіжного модуля, влаштованого в програму займатиме місяць загалом.
	Орієнтація на клієнтів банку	20%	6 місяців

5.4 Розроблення ринкової стратегії проекту

Опишемо цільові групи потенційних споживачів (табл. 14).

Таблиця 5.14 Вибір цільових груп потенційних споживачів

п/п	Опис профілю цільової групи потенційних клієнтів	Готовність споживачів сприйняти продукт	Орієнтовний попит в межах цільової групи (сегменту)	Інтенсивність конкуренції в сегменті	Простота входу у сегмент
	Великі банки	Незначна	За відсутності інших крупних клієнтів - малий	Велика	Важко
	Середні банки	Значна	Великий	Середня	Легше за інші, проте не легко
	Малі банки та ломбарди, інші дрібні позичальники	Залежить від специфіки закладу.	Попит середній	Середня	Середня
<p>Які цільові групи обрано:</p> <p>Система має розроблятися гнучко. Для подальшого розширення на різні сегменти ринку, проте в умовах стартапу інвестиції на початку сильно обмежені тому має сенс працювати з банками середнього розміру, оскільки рівень конкуренції та бюрократії там менший. З плином коштів у проект можна буде розширитися на інші сфери.</p>					

Таблиця 5.15 Визначення базової стратегії розвитку

п/п	Обрана альтернатива розвитку проекту	Стратегія охоплення ринку	Ключові конкурентоспроможні позиції відповідно до обраної альтернативи	Базова стратегія розвитку
	Модель М. Портера	Стратегія зростання	Розвиток продукту	

Наступним кроком є вибір стратегії конкурентної поведінки (табл. 16).

Таблиця 16 Визначення базової стратегії конкурентної поведінки

п/п	Чи є проект «першопрохідцем» на ринку?	Чи буде компанія шукати нових споживачів, або забирати існуючих у конкурентів?	Чи буде компанія копіювати основні характеристики товару конкурента, і які?	Стратегія конкурентної поведінки*
-----	--	--	---	-----------------------------------

	Так	Шукати нових користувачів та після перших успіхів забирати існуючих у конкурентів	Ні	Створити унікальний продукт
--	-----	---	----	-----------------------------

Таблиця 5.17 Визначення стратегії позиціонування

п/п	Вимоги до товару цільової аудиторії	Базова стратегія розвитку	Ключові конкурентоспроможні позиції власного стартап-проекту	Вибір асоціацій, які мають сформувати комплексну позицію власного проекту (три ключових)
	Оптимізація, зниження ризиків та кастомізація	Створити можливість налаштувати коефіцієнт правдивості	Відсутність людського фактору та впливання стереотипів	Оптимізація, зниження ризиків та кастомізація

5.5 Розроблення маркетингової програми стартап-проекту

5.5.1 Першим кроком є формування маркетингової концепції товару, який отримає споживач. Для цього у табл. 18 підсумуємо результати попереднього аналізу конкурентоспроможності товару.

Таблиця 18. Визначення ключових переваг концепції потенційного товару

п/п	Потреба	Вигода, яку пропонує товар	Ключові переваги перед конкурентами (існуючі або такі, що потрібно створити)
	Оцінка платоспроможності клієнта на певну дату	Оптимізація ресурсів, затрачених на оцінку	Швидка оцінка з мінімальним людським втручанням
	Прогноз фінансової стабільності на перспективу	Оптимізація ресурсів, затрачених на прогноз	У конкурентів оцінка на теперішній момент, не на майбутнє.

5.5.2 Надалі розробляється трирівнева маркетингова модель товару: уточнюється ідея продукту та/або послуги, його фізичні складові, особливості процесу його надання (табл. 19).

Таблиця 5.19 Опис трьох рівнів моделі товару

Рівні товару	Сутність та складові		
I. Товар за задумом	Програма, що виконує об'єктивну оцінку фінансової стабільності позичальника. Облік можливих ризиків по кредитних операціях дозволяють банку ефективно управляти кредитними ресурсами й отримувати прибуток.		
II. Товар у реальному виконанні	Властивості/характеристики	М/Нм	Вр/Тх /Тл/Е/Ор
	1. Економічна		
	2. Технічна		
	Якість: стандарти, нормативи, параметри тестування тощо		
	Пакування: Розповсюдження через інтернет		
III. Товар із підкріпленням	До продажу: Мінімальний робочий функціонал, оцінка за регресійною моделлю для банків		
	Після продажу: розширення функціоналу під теперішню та бажану сферу покупців.		
За рахунок чого потенційний товар буде захищено від копіювання: захист коду програми стандартними методами			

Стандартний захист ІТ продукту.

5.5.3 Наступним кроком є визначення цінових меж, якими необхідно керуватись при встановленні ціни на потенційний товар (остаточне визначення ціни відбувається під час фінансово-економічного аналізу проекту), яке передбачає аналіз ціни на товари-аналоги або товари субститути, а також аналіз рівня доходів цільової групи споживачів (табл. 20). Аналіз проводиться експертним методом.

Таблиця 5.20 Визначення меж встановлення ціни

п/п	Рівень цін на товари-замінники	Рівень цін на товари-аналоги	Рівень доходів цільової групи споживачів	Верхня та нижня межі встановлення ціни на товар/послугу
	-	130 000	7 000 000	25 000 – 75 000

5.5.4 Наступним кроком є визначення оптимальної системи збуту, в межах якого приймається рішення (табл. 21):

Таблиця 5.21 Формування системи збуту

п/п	Специфіка поведінки цільових клієнтів	Функції збуту, які має виконувати постачальник товару	Глибина каналу збуту	Оптимальна система збуту
	Бюрократія, між ознайомленням товару та його купівлею рішення має бути прийняте на багатьох рівнях керівництва.	Розмістити на офіційному сайті можливість купити продукт.	Інтернет, SEO	Реклама на банківських конференціях

5.5.5 Останньою складовою маркетингової програми є розроблення концепції маркетингових комунікацій, що спирається на попередньо обрану основу для позиціонування, визначену специфіку поведінки клієнтів (табл. 22).

Таблиця 5.22 Концепція маркетингових комунікацій

п/п	Специфіка поведінки цільових клієнтів	Канали комунікацій, якими користуються цільові клієнти	Ключові позиції, обрані для позиціонування	Завдання рекламного повідомлення	Концепція рекламного звернення
	Ввести рекламну компанію про користь процесу автоматизації аналізу кредитоспроможності	Конференції, мітинги, інтернет	Інтернет та конференції	Донести прибутковість ідеї.	Грошей буде більше

5.6 Висновки

- Світ розвивається в сторону ІТ, є можливість ринкової комерціалізації проекту (чи наявний попит, динаміка ринку, рентабельність роботи на ринку);
- Входити на ринок з готовим інтерфейсом, сайтом та UX буде важче, ніж створити просто прототип. Ще треба вкластися в пошук клієнтів, рекламу і багато іншого. Бар'єр для виходу на великих клієнтів дуже високий також і за рахунок бюрократії останніх.

- Альтернативно можна використовувати крипто-валюти замість звичайних для прогнозу і спробувати зайняти цю нішу.
- Чи є доцільною подальша імплементація проекту? – Так. Можна розробити багато додаткового функціоналу, який має вартість для користувачів.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Грошово - кредитна система України в умовах ринкових перетворень в економіці: Монографія / За ред. д.е.н., проф. О.В. Дзюблюка. – Тернопіль: «Карт - Бланш», 2007. – 308 с.
2. Е. Ф. Жуков, Л. М. Максимова. Банки и банковские операции. [Текст] – М.: ЮНИТИ, 1997. – 417 с.
3. А.М. Мороз, М.І. Савлук, М.Ф. Пудовкіна. *Банківські операції: Підручник*. [Текст] – К.:КНЕУ, 2000. – 384с.
4. Стаття про кредитний ризик [Електронний ресурс]. – 2018. – Режим доступу до ресурсу: <http://ukr.vipreshebnik.ru/entsiklopediya/56-k/3980-kreditnij-rizik.html>
5. Крейнина М.Н. *Финансовое состояние предприятия. Методы оценки*. [Текст] – М.: ИКЦ “Дис”, 1997. – 338с.
6. Ольшаный А.И. *Банковское кредитование: российский и зарубежный опыт*. [Текст] - М:Русская деловая литература, 1998. – 405 с.
7. *Про затвердження Положення про порядок формування та використання резерву для відшкодування можливих втрат за кредитними операціями банків*: Постанова Правління Національного банку України № 279 від 06.07.07 р.
8. *Про банки і банківську діяльність*: Закон України, затв. Верховною Радою України №2121-III від 07.12.00 р., зі змінами та доповненнями.
9. А.М. Єріна. *Статистичне моделювання та прогнозування: Навчальний посібник*. [Текст] -К.: КНЕУ, 2001. – 170 с.
10. Hosmer D.W., Lemeshow S. *Applied Logistic Regression*. [Текст] – A Wiley-Interscience Publication, 2002,

11. Дрейпер, Н., Смит, Г. *Прикладной регрессионный анализ*: В 2-х кн. — М.: Финансы и статистика, 1986.
12. Песаран, М., Слейтер, Л. *Динамическая регрессия: теория и алгоритмы*. [Текст] — М.: Финансы и статистика, 1984. — 487 с.
13. Davis J., Goadrich M. The Relationship Between Precision-Recall and ROC Curves [Текст] / *Proc. Of 23 International Conference on Machine Learning*, Pittsburgh, PA, 2006.
14. Цыплаков А. А. *Некоторые эконометрические методы. Метод максимального правдоподобия в эконометрии*. [Текст] - Новосибирск: ЭФ НГУ, 1997. - 129 с.
15. Калинина В. Н., Соловьев В. И. *Введение в многомерный статистический анализ: Учебное пособие*. [Текст] - М.: ГУУ, 2003. — 66 с.
16. K. Alsabti, S. Ranka, and V. Singh. An Efficient k-means Clustering Algorithm [Текст] / *Proc. First Workshop High Performance Data Mining*, Mar. 1998.- 15 p.
17. E. Forgey. Cluster Analysis of Multivariate Data: Efficiency vs. Interpretability of Classification / *Biometrics*, vol. 21, p. 768, 1965.
18. T. Kanungo, D.M. Mount, N.S. Netanyahu, C. Piatko, R. Silverman, A.Y. Wu. The Analysis of a Simple k-means Clustering Algorithm [Текст] / *Technical Report CAR-TR-937*, Center for Automation Research, Univ. of Maryland, College Park, Jan. 2000. — 11 p.
19. Дубров А. М., Мхитарян В. С., Трошин Л. И. *Многомерные статистические методы: Учебник*. [Текст] — М.: Финансы и статистика, 2000. — 352 с.
20. J. Ross Quinlan. *C4.5: Programs for Machine learning*. [Текст] - Morgan Kaufmann Publishers, 1993.- 543 p.

- 21.К. Шеннон. *Работы по теории информации и кибернетике*. [Текст] - М.: Иностранная литература, 1963. – 829 с.
- 22.С.А. Айвазян, В.С Мхитарян. *Прикладная статистика и основы эконометрии*. [Текст] - М.: Юнити, 1998. – 1022 с.